

Zabal, Anouk; Martin, Silke; Massing, Natascha; Kapidzic, Sanja; Gauly, Britta; Rammstedt, Beatrice
PIAAC Germany 2023. Technical Report

Münster ; New York : Waxmann 2025, 157 S.



Quellenangabe/ Reference:

Zabal, Anouk; Martin, Silke; Massing, Natascha; Kapidzic, Sanja; Gauly, Britta; Rammstedt, Beatrice:
PIAAC Germany 2023. Technical Report. Münster ; New York : Waxmann 2025, 157 S. - URN:
urn:nbn:de:0111-pedocs-351882 - DOI: 10.25656/01:35188; 10.31244/9783818851095

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-351882>

<https://doi.org/10.25656/01:35188>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz:
<http://creativecommons.org/licenses/by/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License:

<http://creativecommons.org/licenses/by/4.0/deed.en> - You may copy, distribute and render this document accessible, make adaptations of this work or its contents accessible to the public as long as you attribute the work in the manner specified by the author or licensor. By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft



Anouk Zabal, Silke Martin,
Natascha Massing, Sanja Kapidzic,
Britta Gauly, Beatrice Rammstedt

PIAAC Germany 2023

Technical Report

WAXMANN



Funded by:



Federal Ministry
for Education, Family Affairs, Senior Citizens,
Women and Youth

Anouk Zabal, Silke Martin, Natascha Massing,
Sanja Kapidzic, Britta Gauly, Beatrice Rammstedt

PIAAC Germany 2023

Technical Report



Waxmann 2025
Münster • New York

Funded by:



Federal Ministry
for Education, Family Affairs, Senior Citizens,
Women and Youth

With support from:



Federal Ministry
of Labour and Social Affairs

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>

Print-ISBN 978-3-8188-0109-0

E-Book ISBN 978-3-8188-5109-5

<https://doi.org/10.31244/9783818851095>

Waxmann Verlag GmbH, 2025
Steinfurter Straße 555, 48159 Münster

www.waxmann.com
info@waxmann.com

Cover Design: Anne Breitenbach, Münster
Typesetting: Roger Stoddart, Münster

This e-book is available under the CC BY 4.0 open access license.



This licence applies only to the original material. All marked third-party content (e.g., illustrations, photos, quotations, etc.) is excluded from the CC licence, and it may be necessary to obtain further permission from the copyright holder for its reuse.

Content

Contributors	9
1 Introduction	11
2 PIAAC Design and Instruments	19
2.1 Background Questionnaire	19
2.1.1 Content of the Background Questionnaire	20
2.1.2 German Adaptations and Extensions	22
2.1.3 Measurement of Education and Education Consultation	26
2.1.4 Changes Between Cycle 1 and Cycle 2	26
2.1.5 Administering the Background Questionnaire	28
2.1.6 Doorstep Interview	29
2.2 Cognitive Assessment	30
2.2.1 PIAAC Skill Domains	30
2.2.2 Item Development	33
2.2.3 Assessment Design	36
2.2.4 Administering the Assessment	40
2.3 Testing the Instruments	41
2.3.1 Testing the Background Questionnaire	41
2.3.2 Testing the Assessment	43
2.4 Instrument Translation	44
2.4.1 Collaboration with Other Countries	45
2.4.2 Translation Staff	45
2.4.3 Translator Training	46
2.4.4 Translation of the Background Questionnaire	46
2.4.5 Translation and Adaptation of the Cognitive Assessment	48
2.4.6 Review After Field Trial	50
3 Preparation for Fieldwork	51
3.1 Sample Design and Selection	51
3.1.1 Target Population, Sampling Frame, and Sample Size	52
3.1.2 Sample Design and Sample Selection in Germany	53
3.2 Outreach	55
3.2.1 Information Material for Target Persons	55
3.2.2 Incentives	57
3.2.3 Invitation Package	58
3.3 Fieldwork Staff	59
3.3.1 Interviewer Selection and Characteristics	59
3.3.2 Interviewer Training	61
3.3.3 Interviewer Manual and Materials	64
3.3.4 Interviewer Remuneration and Incentives	65
3.4 PIAAC Workflow and Case Management System	66
3.5 Hardware and Technical Implementation	69

4	Implementation of Fieldwork	73
4.1	Field Period Organization and Sample Releases	73
4.2	Contacting Rules and Disposition Codes	76
4.2.1	Contacting Rules	76
4.2.2	Disposition Codes and Documentation of Reasons for Nonresponse	77
4.3	Fieldwork Monitoring and Follow-Up Strategies	80
4.3.1	International Fieldwork Monitoring	80
4.3.2	National Fieldwork Monitoring	81
4.3.3	Major Problems Affecting Fieldwork and Procedural Adjustments	82
4.3.4	Refusal Conversion and Reissues	84
4.4	Interview Validation	85
4.4.1	Validation of Interviews	86
4.4.2	Validation of Doorsteps, Breakoffs, and Non-Interviews	89
4.4.3	Monitoring Validation Procedures	90
4.4.4	Evaluation of Audio Recordings	91
4.4.5	Validation Results	93
4.5	Fieldwork Results	94
4.5.1	PIAAC Response Rate and Final Disposition Codes	97
4.5.2	Contacting Results and Behaviour	100
4.5.3	Post-Interview Questions	102
4.5.4	Interviewer Debriefing	103
4.6	Curbside Observations	105
5	Weighting, Nonresponse Bias Analyses, and Variance Estimation	107
5.1	Weighting	107
5.1.1	Unknown Eligibility Adjustment	108
5.1.2	Nonresponse Adjustment	109
5.1.3	Calibration	111
5.2	Extended Nonresponse Bias Analyses	112
5.2.1	Comparison of Estimates Before and After Weighting Adjustments	113
5.2.2	Comparison of Weighted Estimates to External Totals	114
5.2.3	Explained Variation in Outcomes	116
5.2.4	Comparison of Estimates From Alternative Weighting Adjustments	117
5.2.5	Analysis of Variables Collected During Data Collection	118
5.2.6	Level-of-Effort Analysis	119
5.2.7	Calculation of the Range of Potential Bias	120
5.3	Variance Estimation	121
5.4	Design Effect	122
6	Data Management, Data Products, and Overall Data Quality	125
6.1	National Data Management Before Submission of National Database	125
6.1.1	National Data Management Process	125
6.1.2	Coding of Education, Occupation, and Other Variables	128
6.2	Data Processing After Submission of National Database	131
6.3	Data Confidentiality	132

6.4 Data Products, Tools, and Documentation	133
6.4.1 Data Products	133
6.4.2 Data Tools	134
6.4.3 Documentation	135
6.5 Overall Data Quality	136
 Appendix A: Doorstep Interview Administration	 139
Appendix B: PIAAC 2023 Advance Letter	143
Appendix C: Post-Interview Questions	145
References	149
List of Figures	154
List of Tables	155
List of Abbreviations	156

Contributors

The following persons were involved in the implementation of PIAAC 2023 in Germany. We thank everyone for the enormous effort and dedication that went into this challenging project. We are also grateful to the international Consortium for their unwavering support and hard work. And finally, a heartfelt thank you to all our respondents. PIAAC Germany 2023 would not have been possible without them.

German National Centre at GESIS – Leibniz Institute for the Social Sciences

National Project Managers

Prof. Dr. Beatrice Rammstedt
Anouk Zabal

National Team

Dr. Britta Gauly
Dr. Sanja Kapidzic
Silke Martin (*Sampling and Data Manager*)
Dr. Natascha Massing

Dr. Jens Bender (2018–2022)

Dr. Désirée Nießen (2021–2023)

Franziska Ziegler (2023–2025)

Secretariat and Administration: Julia Khorshed, Maria Kreppe-Aygün, Marion Rahnke

Student Assistants: Nina Baumann, Sonja Bibow, Finn Brabänder, Lukas Diehlmann, Viktoria Massar, Ricarda Reiner, Sina Schuck, Nina Trenz, Friederike Voß, Marc Zabel

National Experts: GESIS – Leibniz Institute for the Social Sciences

Dr. Dorothee Behr (*Translation*)

Michael Blohm (*Survey Operations*)

Dr. Christian Bruch (*Sampling, Weighting*)

Dr. Barbara Felderer (*Nonresponse*)

Achim Koch (*Survey Operations*)

Dr. Clemens Lechner (*Social-Emotional Skills*)

PD Dr. Débora Maehler (*Migration*)

Dr. Matthias Sand (*Sampling, Weighting*)

Dr. Silke Schneider (*Education, Sociodemographics*)

Dr. Heike Wirth (*Data Privacy*)

National Experts: External

Prof. Dr. Annelies Blom (*Nonresponse*)

Prof. Dr. Aiso Heinze (*Numeracy*)

Prof. Dr. Corinna Kleinert (*Questionnaire*)

Prof. Dr. Dr. h.c. Joachim Funke (*Adaptive Problem Solving*)

Prof. Dr. Petra Stanat (*Literacy*)

Field Agency

Kantar Public (now Verian)

Günter Steinacker (*Project Director*)

Dr. Jana Brix (*Project Director*)

Sarah Schmidt

Anett Gruber

Wladislaw Kay Neumann

And numerous PIAAC interviewers

National IT Coordination

DIPF | Leibniz Institute for Research and Information in Education

Robert Baumann (*National IT Coordinator*)

Occupation and Industry Coding***IEA Hamburg (International Association for the Evaluation of Educational Achievement)***

Guido Martin

Daniela Specks

*And IEA coders***Translators**

Ulrike Efu Nkong

Ilse Freiburg

Beate Staib

National Scientific Advisory BoardProf. Dr. Corinna Kleinert (*Chair*)

Prof. Dr. Harm Kuper

Prof. Dr. Dr. h.c. Manfred Prenzel

Prof. Dr. Christine Sälzer

Prof. Dr. Heike Solga

Prof. Dr. Mark Trappmann

Prof. Dr. Ludger Wößmann

***Representatives for the Federal Ministry for
Education, Family Affairs, Senior Citizens,
Women and Youth (Former: Federal Ministry
of Education and Research)***

Katharina Koufen

Dr. Susanne Ziemek

***Representatives for the Federal Ministry of
Labour and Social Affairs***

Dr. Svenja Dettner

Marie Ullmann

Annette Vogel

1

Introduction

PIAAC, the *Programme for the International Assessment of Adult Competencies*, is a multi-cycle program initiated by the Organisation for Economic Co-operation and Development (OECD), which measures three foundational skills of the adult population (16–65 years) in a comparable manner across the participating countries: *literacy*, *numeracy*, and *problem solving*. The international design and the national data collections implement state-of-the art standards and guidelines in survey methodology and strive to achieve the highest possible data quality. Two PIAAC cycles have been conducted to date. The first PIAAC cycle took place between 2011 and 2018 and included three rounds of data collection, each involving a different group of participating countries. The first round of the second cycle of PIAAC was completed in 2024 and included 31 participating countries.¹ Germany participated both in the first round of the first cycle of PIAAC, with the main study data collection in 2011/2012 (PIAAC Germany 2012), as well as in the first round of the second cycle, with the data collection in 2022/2023 (PIAAC Germany 2023). This report describes the implementation of PIAAC in Germany in Cycle 2.

PIAAC is steered by the Board of Participating Countries (BPC), which consists of representatives of the funding bodies of the participating countries. The OECD works together with the BPC to make strategic decisions. The OECD contracted an international Consortium for the overall international project management of PIAAC Cycle 2. This Consortium was led by *Educational Testing Service* (ETS, USA; overall management, test development, platform development, psychometrics, analysis, and data products), in cooperation with *cApStAn* (Belgium; linguistic quality control), *GESIS – Leibniz Institute for the Social Sciences* (GESIS, Germany; development of the background questionnaire), the *International Association for the Evaluation of Educational Achievement* (IEA, the Netherlands; data cleaning and database preparation), the *Research Centre for Education and the Labour Market* (ROA, Maastricht University, the Netherlands; development of the background questionnaire), and *Westat* (USA; sample design and selection, weighting, nonresponse bias analyses, survey operations, and quality control). A Technical Advisory Group advised the international Consortium, the OECD, and the BPC on key issues. The development of the PIAAC theoretical frameworks and measurement instruments was shaped by international subject matter expert groups comprised of renowned experts.

In Germany, PIAAC is funded by the Federal Ministry for Education, Family Affairs, Senior Citizens, Women and Youth with the participation of the Federal Ministry of Labour and Social Affairs. As in PIAAC Cycle 1, GESIS was contracted by the Federal Ministry of Education and Research (now: Federal Ministry for Education, Family Affairs, Senior Citizens, Women and Youth) to carry out the national project management for PIAAC Cycle 2 in Germany. The National Centre

¹ The second round of the second cycle was launched in 2025, and the main survey data collection is planned for 2027/2028.

at GESIS was supported and advised by a group of national experts. Strategic decisions were guided by a Scientific Advisory Board. Annual meetings were held with the Scientific Advisory Board and representatives of the ministerial offices and governmental funders. The National Centre subcontracted three national organizations for specific work packages: the commercial field agency *Kantar Public* (now: *Verian*) for the data collection, the *DIPF | Leibniz Institute for Research and Information in Education* (DIPF) for the national information technology coordination, and the *International Association for the Evaluation of Educational Achievement* (IEA Hamburg) for coding activities. Carrying out a study of this scope and complexity requires the collaborative input and extraordinary commitment of many dedicated people. We acknowledge with gratitude all those who contributed to PIAAC in Germany (see the list at the beginning of this report). We would also like to thank the international PIAAC Consortium for their valuable and dedicated support. Finally, a very special thanks goes to the respondents who participated in our survey.

Features of PIAAC Cycle 2

The target population for PIAAC Cycle 2 consisted of adults in the participating countries aged 16 to 65 who lived in private households at the time of data collection – regardless of citizenship, residence status, or proficiency in the national language. Interviewers specifically trained for PIAAC contacted the selected target persons and administered the approximately two-hour PIAAC interview.

The core component of PIAAC, the direct assessment of cognitive skills, was conducted in the country's national language(s) and was designed to measure a broad range of proficiencies in literacy, numeracy, and adaptive problem solving. The assessment tasks reflect everyday life situations and contexts. In Cycle 2, the tasks were administered on a tablet provided by the interviewer. Respondents worked on the tasks by themselves, without any time restrictions, and were supervised by the interviewer.

The cognitive assessment was complemented by a comprehensive background questionnaire, administered as a personal interview by the interviewer. Information was collected on the respondents' sociodemographic characteristics, formal education, further training, employment and occupational situation, as well as the use of skills at work and in everyday life. The analytical power of the PIAAC data stems from the wide range of background information collected through the questionnaire.

To prevent valuable information from being lost due to language barriers, a short doorstep interview was introduced in Cycle 2. This doorstep interview was offered in several languages and self-administered on a tablet.

The development of the design for PIAAC Cycle 2 aimed to strike a balance between maintaining trend with Cycle 1, while introducing innovations and updating the theoretical frameworks and survey instruments to reflect societal and technological changes in the last decade. A set of linking items as well as the core of the measurement for the cognitive domains *literacy* and *numeracy* were maintained in Cycle 2, thus making it possible to compare the data of the two cycles. At the same time, more development work went into improving and extending the cognitive assessment, for example by achieving more precise measurements at lower and higher proficiency levels. In addition, construct coverage was enhanced

through new item development. The third assessment domain, adaptive problem solving, was newly developed for PIAAC Cycle 2. It reflects an increasingly important skill in an era of rapidly accelerating change: the ability to monitor and adapt cognitive strategies in order to deal with dynamic problem situations.

The Cycle 2 background questionnaire was designed to collect comprehensive contextual information, enabling detailed analysis of the relationship between cognitive skills and social and educational variables as well as economic and non-economic outcomes. As far as possible, the Cycle 1 questionnaire was retained in Cycle 2 to allow for trend analyses. However, some changes and innovations were also introduced here, including an updated measurement of education and a new module on social and emotional skills.

The PIAAC interview was administered using specially developed international software on tablet computers that met a set of specifications. This ensured standardized and comparable administrations across participating countries. While the administration of PIAAC Cycle 1 was already computer based – an innovation at the time – the transition to tablets introduced touch functionalities and new possibilities for cognitive item development. It also allowed for more flexible usage. Furthermore, the tablets and more intuitive software design were more accessible and easier to use, even for those individuals who were less tech-savvy. The tablet-based assessment also allowed for fully automated scoring and captured process data.

As in Cycle 1, the international Consortium implemented an elaborate system of quality assurance through comprehensive and mandatory standards and procedures for all phases of the PIAAC survey life cycle. Quality control was carried out through review of national processes, activities, and data throughout the PIAAC workflow. At the national level, countries had their own quality management systems in place.

Field Trial

An important component of quality assurance in PIAAC was a mandatory field trial in preparation for the main study. The field trial had four key objectives: (1) testing the psychometric properties of items, evaluating the functioning of newly developed items, and checking the stability of the item characteristics, (2) evaluating the timing and comparability of items across countries and identifying possible translations errors, (3) assembling the measurement instruments for the main study based on the aforementioned analyses, and (4) functioning as a dress rehearsal for all national procedures, such as sampling and survey operations. The field trial was scheduled to take place from April to June 2020 with a target of 1 500 completed interviews. However, the interviewer training and data collection had to be cancelled at the last minute due to the outbreak of the COVID-19 pandemic. After a period of uncertainty as to how to continue with PIAAC Cycle 2, and following a risk evaluation, the OECD and the BPC decided to revise the project timeline and postpone the field trial to 2021. At the same time, certain standards for the field trial were relaxed to accommodate the different national pandemic contexts and regulations, as even with the one-year delay the field trial was still conducted under restrictive pandemic conditions. For example, the OECD and the international Consortium accepted reduced sample sizes for the field trial and permitted the data collection period to begin flexibly.

A reduced operational field trial was carried out in Germany with various adaptations compared to the original design. Instead of a probability-based sample, a quota sample had to be implemented, with sample persons recruited from interviewers' private networks. Interviewer trainings were held virtually instead of in-person. An experiment to investigate how to maximize response rates had to be cancelled due to the reduced scope of the field trial. Strict COVID-19 health and safety protocols were followed (e.g., masks, hand and screen disinfection, keeping distance), and only vaccinated or immunized interviewers worked on PIAAC. The health of target persons and interviewers was always the highest priority. The German field trial data collection took place from May 31 to July 25, 2021, and realized a sample size of 445 cases.

Main Study

The implementation of the Cycle 2 main study in Germany followed the international PIAAC standards and guidelines and national best practices. A probability-based sample was drawn from national registries, and several measures were put in place to maximize response rates and minimize nonresponse bias. These measures included appointing experienced interviewers, offering attractive interviewer remuneration and extensive in-person training, and developing a tailored, appealing outreach strategy. Comprehensive systems of fieldwork monitoring, interview validation, and quality control were set up and operational both at the field agency and at the National Centre.

In Germany, the PIAAC Cycle 2 main study data collection was carried out by the field agency Kantar Public and took place from September 1, 2022 to May 31, 2023. Although the major restrictions of the COVID-19 pandemic were over at this time, the aftereffects of COVID-19 had a significant negative impact on fieldwork and posed a number of challenges for the data collection – not only in Germany, but in many of the PIAAC countries. Despite the slow and difficult data collection process, 4 731 completed cases were achieved, surpassing the PIAAC requirement (given the national sample design) of 4 500 completed cases. The PIAAC response rate for Germany was 45 %, which is a remarkable achievement given the circumstances. However, this was below the PIAAC standards of 70 % or at least 50 % response rates. Box 1.1 summarizes the key facts about PIAAC Cycle 2 – both from the international and national perspective.

Box 1.1 Key Facts About PIAAC Cycle 2

- Countries Participating in PIAAC Cycle 2
Austria, Belgium (Flanders), Canada, Chile, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, the Netherlands, New Zealand, Norway, Poland, Portugal, Singapore, Slovak Republic, Spain, Sweden, Switzerland, United Kingdom (England), United States
- Assessment Domains
 - literacy, incl. reading components
 - numeracy, incl. numeracy components
 - adaptive problem solving
- Background Questionnaire
 - demographic and background information, education and training, questions related to work, non-economic outcomes, and social and emotional skills
 - in addition to the core questionnaire, each country could add up to 5 extra minutes of national questions; **Germany added seven questions**
 - two international options were offered: (1) additional questions on the work environment, (2) extended social and emotional skills module; **Germany implemented option 2**
- Doorstep Interview
 - very short at-the-door questionnaire for persons with a language barrier
- Interview Administration
 - language of administration: in official country language(s); **Germany: in German**
 - background questionnaire: administered as a computer-assisted personal interview (designed to be 45 minutes; **average duration in Germany: 48 minutes**)
 - cognitive assessment: tablet-based, no timing restrictions, self-administered by respondent, and monitored by interviewer (designed to be 60 minutes; **average duration in Germany: 75 minutes**)
- Target Population
 - adults between 16 and 65 years of age
 - residing in the country during data collection, non-institutionalized
 - regardless of citizenship, residential status or language skills
- Sample
 - probability-based sample, representative of target population in country
 - **Germany: registry-based, two-stage stratified and clustered sampling design with selection of municipalities as primary sampling units at the first stage and selection of individuals at the second stage**
 - number of selected municipalities: 275
 - number of sample points: 320
 - gross sample size: 11 200 target persons
- Target Response Rate
 - goal of 70%, at least 50%
 - for response rates below 70%: nonresponse bias analyses should show no bias or no serious levels of bias; **Germany: some nonresponse bias may exist, PIAAC standard was attained to a sufficient extent**
- Target Minimum Sample Size
 - 4 000 to 5 000 completed cases, depending on the sample design
 - **Germany: 4 500 completed cases**
- Data Collection
 - data collection period foreseen for September 2022 to April 2023, later extended to August 2023
- Data Collection in Germany
 - data collection period: September 2022 to May 2023 (273 days of fieldwork)
 - field agency: Kantar Public (now Verian)
 - number of interviewers: 100
 - number of realized interviews: 4 675
 - number of realized doorstep interviews: 118
 - number of completed cases, total (according to PIAAC definition): 4 731
 - net sample size (cases in final dataset with weights): 4 793
 - achieved response rate (according to PIAAC definition, design weighted): 44.9%

Ethics and Data Privacy

The German National Centre is committed to rigorous professional and ethical standards, as are all sub-contractors. All persons involved in the implementation of PIAAC in Germany or with access to confidential instruments or other confidential PIAAC material were required to sign a confidentiality agreement (national team, sub-contractor staff, experts, etc.). Ethical responsibilities were an explicit part of interviewer training, which covered ethical and professional principles in conducting research, such as being neutral and respectful, the voluntary nature of survey participation, protecting the personal rights of respondents, and ensuring appropriate and safe processing of personal data. An application for ethical review was submitted to the GESIS Ethics Committee and the project plan for PIAAC in Germany was approved.

Appropriate technical and organizational measures and safeguards were put in place to ensure confidentiality, data integrity, and compliance with the General Data Protection Regulation (GDPR) and national data privacy laws. The measures were discussed and agreed upon with the external GESIS Data Privacy Officer and, as appropriate, with the field agency. The specific data privacy information sheet was also drafted and finalized with the GESIS Data Privacy Officer and the field agency. Suitable processes were installed, for example, a formal process for obtaining and documenting the informed consent of the target person. The importance of data security and the confidentiality of responses, data, and all survey material were emphasized in the interviewer manual and interviewer training. The field agency was the first point of contact for target persons with data privacy concerns. The National Centre and the GESIS Data Privacy Officer were informed and consulted by the field agency as circumstances required. Data products were subject to meticulous and systematic screening by the National Centre, and decisions with respect to which data could be released were made together with the national scientific expert for data confidentiality and the external Data Protection Officer. Additional special security measures were installed during the period of the OECD embargo in the months prior to the release of the PIAAC Cycle 2 results.

Report Outline

This technical report provides detailed information on the implementation of PIAAC Cycle 2 in Germany. Each chapter of this report focuses on a specific aspect of the study. *Chapter 2* describes the survey instruments, including their content, administration, development, and translation. *Chapter 3* discusses the preparatory activities required prior to the data collection. This includes topics such as sample design and selection, outreach materials, interviewer recruitment and training, and the PIAAC workflow. *Chapter 4* addresses the implementation of fieldwork, describing survey operation procedures, fieldwork challenges and interventions, monitoring activities, interview validation, and key outcomes such as response rates. *Chapter 5* provides detailed information on weighting procedures and nonresponse bias analyses. Finally, *Chapter 6* presents information on data management, data products and related documentation, as well as a brief evaluation of the final quality of the PIAAC Germany 2023 data.

This report can be regarded as the sequel to the PIAAC Germany 2012 Technical Report from Cycle 1 (Zabal et al., 2014). It also concludes a series of

national publications revolving around PIAAC Germany Cycle 2. These publications include the national report with results for Germany, which presents key findings from PIAAC 2023 (Rammstedt et al., 2024a), as well as a national report focusing on the results for adaptive problem solving (Kapidzic et al., 2025). At the international level, key reference materials include OECD publications such as the report presenting the international results (OECD, 2024a), the International Technical Report (OECD, 2025c), and the Technical Standards and Guidelines (OECD, 2022).²

² These international documents as well as the national fieldwork report produced by the field agency (Kantar Public, 2024) provided valuable information for this technical report. In addition, we acknowledge the use of AI tools to optimize text elements.

2

PIAAC Design and Instruments

Providing high-quality, internationally comparable skill measures for literacy, numeracy, and adaptive problem solving is a central objective of PIAAC Cycle 2. Cognitive skills were measured using a direct assessment, which was developed by the international Consortium together with domain expert groups and with contributions from countries (see Section 2.2). The analytical potential of the PIAAC data stems from the rich information collected via a background questionnaire (see Section 2.1). The development of the background questionnaire was also the responsibility of the international Consortium, in consultation with experts from different fields relevant to the questionnaire.

The development of survey instruments is an elaborate process. Beyond their theoretical underpinnings, it is crucial that the instruments are culturally appropriate and comparable across the participating countries. To this end, instrument development in PIAAC Cycle 2 underwent many iterations, including feedback from the countries with respect to cross-cultural and cross-linguistic issues. National Centres were responsible for translating and adapting the source instruments through a state-of-the-art translation process (see Section 2.4). The translations underwent a strict verification process carried out by the international Consortium. All instrument modules were extensively tested (see Section 2.3). The PIAAC field trial, which was mandatory for all countries, was the final critical component of instrument validation and item selection.

The PIAAC interview consists of two components. First, the background questionnaire was administered by an interviewer as a computer-assisted personal interview (CAPI). The subsequent cognitive assessment was self-administered on a tablet but monitored by the interviewer. Interviews were conducted in the official language(s) of each respective country. In Germany, the interviews were administered in German. Interviews were usually carried out at the respondent's home.

2.1 Background Questionnaire

The development of the PIAAC Cycle 2 background questionnaire was based on a conceptual framework, which provides a theoretical rationale for including different constructs and defines and elaborates the underlying concepts and how they are measured, as well as their analytical potential (OECD, 2018). The objectives of PIAAC, according to the conceptual framework, are twofold. First, to measure the availability of essential cognitive, non-cognitive, and other key skills for successful participation in work and society across different countries. And second, to explore how these skills are related to important economic and social outcomes and to investigate which factors influence the development, maintenance, and decline of skills (OECD, 2018). The aim of the background questionnaire is to provide relevant information in order to fulfil these objectives.

The PIAAC Cycle 2 background questionnaire and its conceptual framework are largely based on their Cycle 1 counterparts (OECD, 2013). However, the background questionnaire was carefully reviewed, revised, and extended before its implementation in Cycle 2. The aim was to keep most of the questions the same in order to maintain a trend between Cycles 1 and 2. The main goal of the revision for Cycle 2 in comparison to Cycle 1 was to adapt the questionnaire to better represent changes in society, especially due to technological changes in the last decade, updating it to new international standards, and introducing new concepts that had not been previously measured (OECD, 2024c). The revision process for Cycle 2 was advised and monitored by the Background Questionnaire Expert Group – consisting of internationally renowned experts from different disciplines, especially from the areas of education and the labour market.

Part of the development phase of the background questionnaire for Cycle 2 was its validation by the international Consortium after the field trial. For the validation, the background questionnaire was tested, and data was analyzed to ensure that questions and routing functioned correctly, to check the length of the background questionnaire, and to assess item distributions and nonresponse. Following the analyses, small changes were implemented, and the background questionnaire was finalized for the main study.

2.1.1 Content of the Background Questionnaire

The aim of the background questionnaire was to gather rich information in order to assess how skills are distributed across different groups and to answer questions related to the development, maintenance, and decline of skills. At the same time, keeping respondents' burden in mind was another important objective when developing the background questionnaire. Finding the right balance between asking relevant questions and length thus also influenced the development. The international background questionnaire was designed to take 45 minutes, with the option of adding five additional minutes for country-specific questions (see Section 2.1.2).

The background questionnaire provides detailed demographic information such as age, gender,³ country of birth, citizenship, and languages. The questionnaire includes various questions related to formal education, such as highest educational attainment, current participation in formal education, and uncompleted education, as well as field of study. Furthermore, there is an additional question related to educational pathways which covers educational qualifications achieved prior to the highest qualification. In addition to formal education, the background questionnaire includes detailed measures on participation in non-formal training over the last 12 months. These questions cover topics such as incidence, intensity, content, and costs of training, as well as barriers to participation.

The background questionnaire also collects limited information about the respondent's work history and, where applicable, detailed information about their current or last job. This includes key descriptive indicators of the job held (e.g., occupation, industry sector, working hours, and job satisfaction), as well as the

3 In PIAAC, the term *gender* is used although the concept measured is sex.

level of education required for the job. The questionnaire includes various skill-use scales related to work which are connected to the competencies measured in the direct assessment (literacy, numeracy, and adaptive problem solving), but which also extend to other areas, such as ICT⁴ practices. In addition, several questions relate to the working environment.

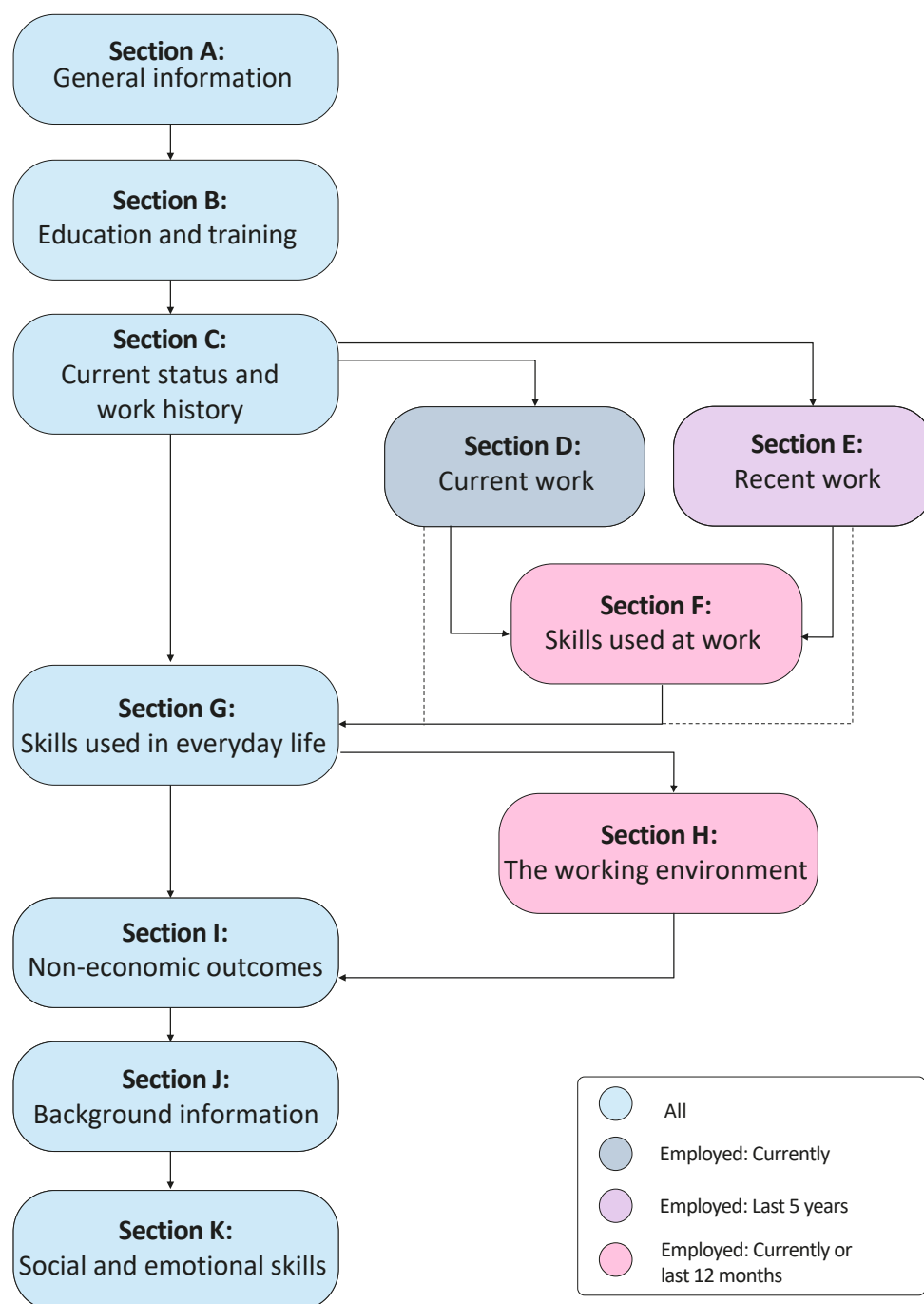
Furthermore, questions were asked about skills used at home as well as non-economic outcomes, such as volunteering, social trust, subjective health, and – in more detail – social and emotional skills (non-cognitive skills).

Finally, there were several questions on the background of respondents, including details about partners, children, as well as some information about the respondent's situation at age 14 (e.g., parental occupation and education, siblings, and place of residence).

In addition to the content illustrated above, which can be described as the *core* background questionnaire, the background questionnaire included two international options: an additional set of questions on the work environment and an additional set of questions in the social and emotional skills module. Countries could choose to implement one, both or none of these options. Germany selected the international option for social and emotional skills. The option includes the full 30-item version of the Big-Five inventory (BFI-S) rather than the 15-item version (BFI-XS; for both scales see Soto & John, 2017) that was part of the international core background questionnaire. Each country also had the opportunity to add a limited number of questions that addressed national priorities. A maximum of five additional minutes were allowed for these questions (see Section 2.1.2 for more information on the German national extensions).

The background questionnaire is divided into different sections, each with a different focus. Not all respondents were asked questions from every section (e.g., the section on employment was not relevant for unemployed respondents). Figure 2.1 shows an overview of the different sections and the overall routing.

4 ICT = Information and Communications Technology

Figure 2.1: Overall Flow of the PIAAC Cycle 2 Background Questionnaire

2.1.2 German Adaptations and Extensions

PIAAC was designed to generate comparable data. The process of achieving comparable data is called harmonization (Hoffmeyer-Zlotnik & Wolf, 2003; Wolf et al., 2016). This harmonization process can either be achieved by input or by output harmonization. Input harmonization refers to the process of designing survey questions so that they are conceptually equivalent and comparable across different

countries and can be translated directly from the international source. During the translation process, it may be necessary to change the wording of a question or response option to ensure that the intended concept is accurately reflected (see Section 2.4). This linguistic adaptation is not what is meant by adaptation here. Output harmonization refers to structural changes to the questionnaire to ensure that it measures the same concept. These changes are then recoded into internationally comparable response categories. In PIAAC, the output harmonization was *ex-ante*, which means that the harmonization was planned during the development of the background questionnaire. Thus, in the following, *adaptations* refer only to structural adaptations that impact the structure of the background questionnaire.

Some adaptations were internationally required, for example, adaptations to the questions on formal education, questions referencing a list of countries or languages, and questions on occupation. A few other questions were flagged as potential adaptations and needed to be reviewed by countries. All national adaptations required approval by the international Consortium. They were documented and justified by countries in an Excel file (Background Questionnaire Adaptation Spreadsheet) and subsequently reviewed by the international Consortium as a part of the sign-off process. Box 2.1 lists the most important adaptations in the German background questionnaire.

Box 2.1: German Adaptations**a) Educational Qualifications**

There were various questions relating to formal educational qualifications – one of the key background variables in PIAAC – that required adaptation to national systems. The aim was to ask about educational qualifications in the national systems and then to code them into the *International Standard Classification of Education 2011* (ISCED 2011; UNESCO Institute for Statistics, 2012). The education questions in the background questionnaire referred to a variety of contexts, such as highest qualification, current education, and uncompleted educational qualification, as well as highest qualification of mother and father. As a new question in Cycle 2, a question on educational pathways was included in the background questionnaire. Furthermore, questions on the qualifications required for the respondent's current job were used to measure education mismatch.

Due to Germany's educational system, two main questions were used to gather information on educational qualifications instead of one question in the international background questionnaire. The first question focused on general education (*allgemeinbildender Schulabschluss*), while the second question covered professional or university education (*beruflicher Ausbildungsabschluss/Hochschulabschluss*). Depending on the answers given in the two main questions, follow-up questions were asked, for example, in order to identify how long individuals had attended school. The response categories (i.e., different qualifications) for these questions were chosen in consultation with national educational experts and underwent an education consultation process (see Section 2.1.3) with the international Consortium in order to determine the best response categories. The fact that two questions and not only one were needed to obtain the required information on educational qualifications also affected the routing of the German background questionnaire.

In contrast to the detailed levels of respondent's education, parental education was only measured in broad levels (differentiating between low/ISCED 0–2; medium/ISCED 3–4; and high/ISCED 5–8) in the international background questionnaire. However, the German adaptation of these questions collected information on parents' educational attainment in two separate questions – one for general education and one for professional or university education – allowing for more detailed distinctions.

b) Occupation

Each country was required to adapt the questions on respondents' and parental occupation in an optimal way for their country, in order to obtain enough detail to be able to code the answers into the *International Standard Classification of Occupations 2008* (ISCO-08; International Labour Office, 2025). In Germany, questions about occupation were asked in two open-answer questions, using the same wording as in Cycle 1 in order to maintain consistency between both cycles. In Cycle 1, the wording was based on questions from national surveys and the German demographic standards (Statistisches Bundesamt, 2004). Although the German demographic standards had been updated after Cycle 1, the questions on occupation remained the same (Statistisches Bundesamt, 2016). In addition to the wording in the demographic standards, the first question includes examples of potential responses adapted to the German context (e.g., *Please give a FULL description, i.e. [...] instead of merely stating "worker", give the exact title, for example "machine fitter"*).

c) Industry

The questions on industry were designed to be coded into the *International Standard Industrial Classification of all Economic Activities Rev. 4* (ISIC; United Nations, 2008). In Germany, there were two open-answer questions, which remained the same as in Cycle 1.

d) Salary

The questions on salary (or wages/earnings) were designed to minimize nonresponse. First, respondents were asked in which intervals they wanted to report their salary (i.e., per hour, day, week, two weeks, month, year, or piece rate). They were then asked to provide their salary (for their selected interval) in an open format. If they were unwilling or unable to provide an answer, they were asked to choose one of six salary ranges. These salary ranges were adapted to reflect the distributions in each country and were defined by the average gross salary/wage. The categories were derived by the international Consortium for each country and reviewed by the National Centres. In Germany, the figures from Cycle 1 were taken as a basis. After this, they were adjusted for Cycle 2, using OECD data on the expected increase in salaries/wages since Cycle 1.

e) Countries/Languages

Questions related to countries and languages, such as country of birth and language most often spoken at home, were adapted using a combination of closed and open-answer questions. Respondents were provided with a showcard listing the most common countries or languages. If the respondent mentioned another country or language (response category "other"), a free-text entry follow-up question was asked. The list of most common countries listed in the main study background questionnaire was derived from the German Microcensus 2020 data (Statistisches Bundesamt, 2021). The list of languages was based on the language spoken most frequently at home from the 2020 Microcensus data (Statistisches Bundesamt, 2021).

f) Field of Study

There were several questions in the background questionnaire asking about the field of study (also called main area or emphasis of study) for highest qualification, current education or pathways questions. In addition to the list of fields of studies, examples of possible qualifications were provided for each field. Both the list of fields of study and the corresponding examples had to be adapted to fit the national context.

g) Other Adaptations

There were several additional questions with nationally-specific response categories. For example, the categories of the question related to the type of employment contract held by the respondent were adapted to include the full set of national employment contracts.

As mentioned above, each country could add an additional five minutes to ask questions that focused on national issues and interests (national extensions). In Germany, a total of seven questions were added as national extensions for the main study background questionnaire of Cycle 2.⁵ Three of them had already been implemented in Cycle 1:

- place of residence before the fall of the Berlin Wall in 1989
- identification of persons in active/passive retirement (important for questions related to employment)
- respondent's citizenship

Additionally, some new national extensions were introduced for PIAAC Cycle 2 to allow further analyses (with a focus on migration):

- reasons for immigration
- identification of resettlers (*Aussiedler*)
- self-assessed German language skills (non-native speakers)
- federal state in which school was attended at age 14

Questions that were adapted are identified by an additional country code (in Germany: DE) included in the variable name in the background questionnaire and in the PIAAC data. Extensions are identified by the country code and an additional X (in Germany: DEX).

⁵ In the background questionnaire for the field trial, there were more national extensions. However, as the analysis of data from the field trial showed that the German background questionnaire exceeded the intended 50 minutes, some national extensions were deleted after the field trial.

2.1.3 Measurement of Education and Education Consultation

In PIAAC Cycle 2, formal education was coded into a coding scheme based on the International Standard Classification of Education (ISCED) 2011 (UNESCO Institute for Statistics, 2012). This classification is used to classify and compare educational systems and qualifications across different countries. The process of deriving the lists of educational qualifications for the international variables was different from the process in PIAAC Cycle 1. Additionally, the coding scheme used in Cycle 1 was ISCED 1997 (UNESCO Institute for Statistics, 2025) instead of ISCED 2011. In Cycle 2, there was a comprehensive consultation process between National Centres and the international Consortium. The aim of the education consultation process was to create a correspondence table between the national categories and the coding scheme used for PIAAC. The basis for this correspondence table were the official ISCED mappings from Eurostat (UNESCO Institute for Statistics, 2025). Initially, the National Centre in Germany reviewed the official mappings and added suggestions to best map national qualifications to the coding scheme used in PIAAC. Once agreement had been reached after an iterative process between the National Centre and the international Consortium, the correspondence table was finalized. Based on this correspondence table, different lists of formal educational qualifications were derived, which were used as categories for each question. Although most of the categories were the same for each question, there were differences between some of the questions. For example, the list of current qualifications did not include outdated categories (e.g., “Volksschulabschluss”), because it was no longer possible to obtain these qualifications. However, they were included in the list for highest qualification because they were relevant for older respondents.

2.1.4 Changes Between Cycle 1 and Cycle 2

Although the background questionnaire is largely based on Cycle 1, there are some differences. Items that were included in both cycles are considered to be either *strict trend* if they have not been changed in Cycle 2 or *soft trend* if the concept has remained the same, but the item operationalization was modified in any way. Additionally, some new items were introduced for Cycle 2, and some items from Cycle 1 were deleted. Furthermore, there were changes to the question order in the questionnaire. This section will focus on the most important changes.

The majority of changes to the background questionnaire between Cycle 1 and Cycle 2 were international changes that affected all countries (see Jacobs et al., 2025). In addition, there were a few national changes to the German background questionnaire that were adjustments of question wording to reflect recent changes in society. All changes between Cycle 1 and Cycle 2 were documented in an Excel file.

The most important international modification relates to changes in the measurement of education. First, the updated ISCED 2011 was implemented in PIAAC Cycle 2 as the international classification for education, instead of the 1997 version of ISCED (ISCED 1997) used in Cycle 1. This entails several changes: ISCED 2011 is more complex, and the main levels of education have been divided into sev-

eral categories to better distinguish different educational qualifications. In Cycle 2, it is possible to derive the orientation of qualifications as vocational or general directly, whereas in Cycle 1 this was only possible ex-post. The coding scheme in PIAAC was extended from ISCED 2011, introducing a distinction of vocational qualifications at ISCED levels 6 and 7 (in addition to professional and academic qualifications). This does not affect all countries, but it is relevant for Germany because there are vocationally oriented qualifications at these ISCED levels. For more detailed information on the international changes, see Huijts et al. (2025) and Schneider (2025). As a consequence of classification changes in ISCED 2011, some of the national qualifications are classified at a different educational level than in ISCED 1997. In Germany, this is the case for “Schule des Gesundheitswesens, 2-3-jährig” which was classified as a high level of education in ISCED 1997 and is classified as medium level of education in ISCED 2011. The changes between Cycle 1 and Cycle 2 in ISCED do not allow direct comparisons between ISCED variables. However, the PIAAC datasets include derived variables that have been harmonized to enable comparisons between Cycle 1 and Cycle 2.

There were also some changes to the national measurement of education: The list of qualifications provided to the respondent was more detailed in Cycle 2 than in Cycle 1. This was done to facilitate the response process, making different qualifications more salient, and distinguishing different qualifications at different institutions. For example, “school for medical assistants, 1-year course” (“Schule des Gesundheitswesens, 1-jährig”) was asked separately instead of in a combined category together with completed vocational training (“beruflich-schulische Ausbildung”) as in Cycle 1. Furthermore, “Diplom” and “Master” at the University of Applied Sciences (“Fachhochschule”) were asked in a combined category in Cycle 1 and were both classified as master level. They were separated for Cycle 2, and a “Diplom” obtained at a University of Applied Sciences was classified as bachelor level and a “Master” as master level.

In Cycle 1, there was a separate “Foreign qualification” category, with the possibility of specifying the name of the foreign qualification. These open-text responses had to be coded into the national categories, which proved to be somewhat problematic. In Cycle 2, respondents were asked to map their foreign qualifications directly into the national education categories.

As an innovation, a new question on educational pathway was introduced for Cycle 2. It measures educational qualifications which were completed before the highest qualification was attained and allows for an analysis of educational trajectories. This is especially relevant for countries with tracked systems such as Germany.

The subsection on non-formal training was revised substantially for Cycle 2. Although most of the concepts remain the same (e.g., recent training volume, whether training was job-related, training costs, motivation for and barriers to training), the operationalization was different, so that the questions on non-formal training are not directly comparable between Cycle 1 and Cycle 2. Additionally, new concepts were introduced, for example training content for one training activity, as well as mode of training (in-person or via distance learning methods).

Several questions related to unemployment were added, such as reasons for stopping one’s last job, unemployment benefits and total time out of work. Changes were also made to job-related questions, such as new position at cur-

rent employer and change of tasks. Items on social media were added in the reading and writing scales. The scale measuring numeracy skill use was adapted. The definition of using a computer was expanded to include smartphones and tablets. Finally, a question measuring life satisfaction was added.

As another innovation in the Cycle 2 background questionnaire, a section with items measuring non-cognitive skills – also called social and emotional skills – was included. This new content was added in Cycle 2 due to evidence showing that non-cognitive skills are related to life outcomes, such as educational attainment and income (Lechner et al., 2019).

2.1.5 Administering the Background Questionnaire

The background questionnaire was administered as a computer-assisted personal interview on a convertible tablet computer⁶ (in the computer mode). All routings were implemented in the international PIAAC software, and interviewers followed the instructions on the screen. To ensure a standardized administration across all interviewers and countries, interviewers were extensively trained on protocols for the administration of the background questionnaire as well as on specific content and questions (e.g., the education module). The questions were read out in verbatim. Showcards were used for several questions to facilitate the response process, principally when there were many response categories or when the showcard included examples. The showcards were handed to the respondent in the form of a booklet at the beginning of the interview, and interviewers directed respondents to the specific showcards as needed. When no showcard was foreseen, interviewers read out the response categories in full (if available) or noted responses to open-answer questions. Interviewers were trained to listen actively and probe appropriately, which was especially relevant before entering responses to open-ended questions into the tablet computer. Interviewer instructions and additional information under a help function could also be consulted if needed.

Interviewers did their best to create a good, respectful, and neutral atmosphere for the administration of the background questionnaire. If feasible, they tried to avoid the presence of others in the room. This is important in order to make sure that the respondent is free to respond without pressure.

In Germany, the mean duration of the background questionnaire administration in the main study was 48.5 minutes (median 47.1 minutes), with a minimum length of 18 minutes and a maximum length of 120 minutes.

The PIAAC survey is designed to address all types of individuals, including those with limited or no proficiency in the official country language(s). For target persons with language barriers, the procedure of choice was to find an interpreter for the background questionnaire, preferably a professional interpreter, and have the respondent at least attempt the cognitive assessment. However, if no professional interpreter was available, a family member could be asked to provide translation support. Use of an interpreter (professional or family member) had to be documented in the post-interview questions (see Section 4.5.3). In Germany, there are 11 documented cases where an interpreter was used.

6 PIAAC tablet.

Even though the PIAAC standards explicitly allowed an interpreter during the background questionnaire administration, this was not allowed during the direct assessment. If it was not possible to find an interpreter for the background questionnaire for target persons with language barriers, interviewers could conduct the doorstep interview.

2.1.6 Doorstep Interview

In PIAAC Cycle 1, if no interpreter was available, interviewers were asked to collect age and gender for target persons with language barriers (so-called literacy-related nonrespondents). The intention was to use this basic demographic information to impute proficiency estimates (plausible values) for these persons. However, this did not work out for methodological reasons (OECD, 2013). In PIAAC Cycle 2, the international Consortium sought to address this issue by introducing the *doorstep interview*, which is a very short version of the background questionnaire designed for persons with a language barrier (Jacobs et al., 2025). This instrument is a specific form of a nonresponse questionnaire.

The doorstep interview was only administered under specific conditions: First, the target person did not speak/understand the language of the PIAAC interview (German, in our case). Second, the interviewer was unable to find an interpreter or another multilingual person (e.g., a family member) to assist with the translation of the background questionnaire. Third, all efforts by the interviewer to conduct a regular interview with the target person had failed.

The doorstep interview was completed by the target person directly at the doorstep using the interviewer's PIAAC tablet. They answered six central questions (Jacobs et al., 2025):

- gender
- age group
- number of years of schooling completed (broad categories)
- subjective status (including employment status)
- born in country
- number of years residing in country

Most of the responses to these doorstep questions could be directly compared to the corresponding questions in the background questionnaire. Only the responses to years of schooling had no direct correspondence.

The international Consortium developed and centrally translated the doorstep interview questions and the administration material for interviewers into different languages. Countries could decide how many and which languages they wanted to include in their doorstep instrument based on the composition of their migrant population. In Germany, both the questions and the accompanying material were available in 18 foreign languages.⁷ The selection of languages was based on the following rationale: (a) including languages spoken by larger migrant population

⁷ Languages offered for the doorstep interview in Germany: Amharic, Arabic, Chinese (simplified), Chinese (traditional), English, Farsi, French, Italian, Croatian, Polish, Portuguese, Romanian, Russian, Somali, Spanish, Turkish, Ukrainian, and Urdu.

groups in Germany (e.g., Turkish, Russian, Polish) and (b) including widely spoken global languages (e.g., English, French, Chinese).

Interviewers were equipped with a *doorstep booklet* containing three distinct types of cards (in 18 languages), each serving a specific function. These cards were used to identify target persons with language barriers and to facilitate basic communication at the doorstep. The administration of the doorstep interview is described in Appendix A.

The challenging communication situation at the doorstep with persons who were not proficient in German posed difficulties for interviewers. In some cases, it was not possible to fully follow the intended steps of the doorstep interview workflow. In many cases, target persons with language barriers declined to participate immediately, and a nonresponse code due to language barriers was assigned. At the same time, it was important to ensure that interviewers did not use the doorstep interview as a shortcut for a full interview. It was therefore crucial to monitor and validate doorstep interviews. Despite these challenges, the doorstep interview proved to be a suitable instrument for collecting very basic, yet valuable, information from target persons with language barriers. Doorstep cases were regarded as completed cases in PIAAC 2023. They are therefore included in the net sample, and proficiency estimates (plausible values) were imputed for this subgroup.

2.2 Cognitive Assessment

This section provides a description of the direct assessment of literacy, numeracy, and adaptive problem solving, including an overview of the skill domains, the instruments, and the adaptive assessment design for PIAAC Cycle 2. More detailed information on the assessment instrument and design for the field trial and main study is provided in the International Technical Report (Ali & Robin, 2025; Lennon & Halderman, 2025). The section concludes with a brief overview of the assessment administration in Germany.

2.2.1 PIAAC Skill Domains

The second cycle of PIAAC measures three skill domains: literacy, numeracy, and adaptive problem solving. Theoretical frameworks were developed for each domain by an international group of experts in the specific field (Greiff et al., 2021; OECD, 2021; Rouet et al., 2021; Tout et al., 2021). The assessment frameworks define the measurement constructs and describe core dimensions of the domains. These dimensions and their elaborations are essential both for item construction as well as for the description of proficiency levels. The literacy and numeracy frameworks were based on the Cycle 1 frameworks (Gal et al., 2009; Jones et al., 2009), but with some extensions, notably to incorporate the increasing importance of digital environments and tools, and to improve differentiation at the lower end of the proficiency scales (Lennon & Halderman, 2025). In turn, the Cycle 1 frameworks evolved from the two predecessors of PIAAC: the Adult Literacy and Life Skills Survey (ALL, 2002–2008; Murray et al., 2005; OECD & Statistics Canada, 2005, 2011) and the International Adult Literacy Survey (IALS, 1994–1998; OECD &

Statistics Canada, 2000). The framework for adaptive problem solving was newly developed for PIAAC Cycle 2.

To support item development and to ensure that the assessment instrument covers the constructs appropriately, expert groups specified coverage targets for the different dimensions and identified factors that drive item difficulty. The item pools for literacy and numeracy consisted of new assessment items and selected linking items. The latter are items from PIAAC Cycle 1, which were adapted for the new tablet-based implementation in Cycle 2 and allow the proficiency scales to be psychometrically linked. Although the cognitive assessment was computer-based per default in Cycle 1 (with paper-based items as an alternative),⁸ the software and functionalities differed from those in the tablet-based software, requiring the items to be updated. At the same time, the Cycle 2 technical environment enabled a wider range of tasks, task complexity, and response formats. As in Cycle 1, item development focused on authentic tasks, embedded in everyday contexts, and appropriate for adults aged 16–65 with very heterogeneous cultural and personal backgrounds.

After briefly describing the domain definitions and core dimensions for PIAAC Cycle 2, a succinct overview of item development, item functionalities, and the assessment design are provided (for more details on the theoretical frameworks, see OECD, 2021; for more details on the cognitive items and assessment design, see Ali & Robin, 2025; Lennon & Halderman, 2025).

Literacy

Literacy is a crucial skill for participation in society. In PIAAC Cycle 2, literacy is defined as “*accessing, understanding, evaluating and reflecting on written texts in order to achieve one’s goals, to develop one’s knowledge and potential and to participate in society*” (Rouet et al., 2021, p. 42). The literacy framework defines the following core dimensions of literacy (Rouet et al., 2021): (a) *cognitive processes*, (b) *content*, and (c) *context*. Three cognitive processes are central to the definition of literacy: accessing, understanding, and evaluating. More specifically, readers actively search for texts and look for relevant information within them. Furthermore, they strive to understand the meaning of the text and evaluate it, for example, in order to assess whether it serves their reading goal, and whether the source is credible. The content dimension encompasses different text types (e.g., descriptions, narrations, instructions), various text formats, text organization (e.g., layout and content representation), and whether the text comes from a single source or multiple sources. Finally, the literacy domain takes into account several contexts, such as personal, work, or social contexts (Rouet et al., 2021). The dimensions enable assessment items to be constructed for different proficiency levels and descriptions of abilities at different proficiency levels to be formulated. In Cycle 2, the literacy framework was updated to address changes in society, especially the continuing digitalization of many aspects of communication and information exchange and was extended to specifically encompass complex digital information environments.

⁸ The option of completing the assessment on paper was offered to respondents who lacked basic computer skills or who were unwilling to complete the assessment on the computer.

Reading Components

The assessment of reading components was an innovation in Cycle 1, aiming to provide more detailed measures of low literacy skills (Sabatini & Bruce, 2009). In Cycle 1, the reading components were administered on paper only; interviewers read out the instructions and registered timing using a stopwatch. However, the reading component results were not integrated into the Cycle 1 literacy scale. For Cycle 2, two of the reading component tasks⁹ were implemented and transferred to a tablet-based measurement (Rouet et al., 2021): (a) *sentence processing*, requiring respondents to assess whether a sentence makes sense and (b) *passage comprehension*, requiring respondents to select the correct alternative out of two words within a text. The reading components were fully integrated in the assessment workflow and completely self-administered; performance and timing data were automatically registered. There was no time limit for the reading components in either cycle, although respondents were instructed to work as quickly as possible. In Cycle 2, the reading components were an integral part of the assessment, and they form part of the literacy scale. The reading components for Cycle 2 consisted of Cycle 1 items as well as items from PISA¹⁰ 2018 (Ali & Robin, 2025).

Numeracy

In PIAAC Cycle 2, numeracy is defined as “*accessing, using and reasoning critically with mathematical content, information and ideas represented in multiple ways in order to engage in and manage the mathematical demands of a range of situations in adult life.*” (Tout et al., 2021, p. 93). Similar to the literacy framework, the numeracy framework defines several core dimensions (Tout et al., 2021, pp. 94–95): (a) *cognitive processes*, (b) *mathematical content*, (c) *representations*, and (d) *context*. Three cognitive processes are central to the numeracy framework: accessing and assessing situations mathematically, acting on and using mathematics, and evaluating, critically reflecting, and making judgements. The framework distinguishes between four different types of mathematical content: quantity and numbers, space and shape, relationship and change, and data and chance. Furthermore, the framework defines four different representations of mathematical content: text or symbols, images of physical objects, structured information, and dynamic applications. Similar to texts in the literacy framework, mathematical information can be encountered in various contexts: personal, work, and social context. Compared to Cycle 1, the numeracy framework in Cycle 2 puts more focus on critical reasoning and reflection as well as on diversity of representation, specifically including digital representations and applications.

Numeracy Components

An innovation of the second cycle of PIAAC was the implementation of numeracy components. Similar to the reading components, numeracy components were developed to allow for more detailed measurement and differentiation at the lower end of the numeracy proficiency scale. The numeracy components focus on basic

⁹ In Cycle 1, the reading components included a task on word meaning which was not included in Cycle 2.

¹⁰ Programme for International Student Assessment

number sense, which “relates to the sense of quantities and the sense of how numbers represent quantities” (Tout et al., 2021, p. 135). Numeracy components included two sets of items (Tout et al., 2021): (a) items requiring a *quantity selection*, that is, asking for an answer to the question “how many?” and (b) items on *relative magnitudes*, that is, asking for an answer to the question “the biggest [number]?”. The assessment of the numeracy components was tablet-based and self-administered; performance and paradata were automatically registered. Respondents were asked to respond as quickly as possible, but no time limit was imposed. The numeracy components were an integral part of the numeracy assessment and the numeracy proficiency scale.

Adaptive Problem Solving

In today’s world, characterized by rapid change, rising complexity and ubiquitous digitalisation, the ability to adapt flexibly to change is an essential skill. The domain of adaptive problem solving was newly developed in PIAAC Cycle 2 to account for the increasing relevance of this skill in today’s world (for details see Greiff et al., 2021; Lennon & Halderman, 2025). Adaptive problem solving “involves the capacity to achieve one’s goals in a dynamic situation, in which a method for solution is not immediately available. It requires engaging in cognitive and metacognitive processes to define the problem, search for information, and apply a solution in a variety of information environments and contexts” (Greiff et al., 2021, p. 159). The framework identifies *cognitive and metacognitive processes* as core dimensions of adaptive problem solving. Cognitive processes encompass organizing pieces of information and relating them to each other and evaluating which piece of information is relevant in the process of attaining one’s goal. Metacognitive processes encompass defining one’s goal in the first place and reflecting on the process of reaching it. Cognitive and metacognitive processes are required in all three stages of problem solving, namely defining the problem, searching for information, and applying a solution (Greiff et al., 2021, p. 161). Furthermore, adaptive problem-solving tasks can be described according to several dimensions: (a) problem configuration (e.g., number of parallel tasks), (b) dynamics of the situation (e.g., frequency of change), (c) features of the environment (e.g., number of information sources), (d) information environment (e.g., physical or digital), and (e) problem contexts (e.g., personal, work, social).

2.2.2 Item Development

During the development of the cognitive instruments for Cycle 2, several important technical requirements had to be considered. First, it was necessary to maintain comparability over time, thus Cycle 2 needed to include trend items from Cycle 1. Second, automatic scoring of all items was necessary to allow for adaptive testing (see Section 2.2.3). Therefore, all items needed to be checked and, if necessary, adapted to allow for the scoring process to happen automatically within the system, without the necessity of human scorers processing the data. Finally, response options needed to be optimized for tablet delivery to minimize digital skill demands (Lennon & Halderman, 2025).

The assessment instrument consisted of both newly developed items for PIAAC Cycle 2 and linking items from PIAAC Cycle 1. The adaptive assessment design required 75-80 items each for the literacy and numeracy domains. Approximately 30 linking items per domain were needed to enable the Cycle 2 and Cycle 1 scales to be linked. A pool of 65 items was necessary for adaptive problem solving.

For literacy, 67 new items were developed and tested in the field trial together with 30 linking items. For numeracy, 64 new items were developed and tested in the field trial together with 32 linking items. For adaptive problem solving, 65 new items were developed and tested. The international Consortium analyzed the field trial data, reviewed the psychometric parameters, and evaluated the quality and comparability of the items. Table 2.1 shows the number of items that were selected for the main study. Locator items in the literacy and numeracy item pools included 14 new literacy and 14 new numeracy locator items, as well as two linking items each (the locator test is a module with a routing function in the cognitive assessment, see Section 2.2.3). Eight of these items were selected for each domain in the main study. The number of locator items was increased in comparison to PIAAC Cycle 1 in order to allow for a more accurate measurement of low literacy and numeracy skills (Ali & Robin, 2025).

Table 2.1: Number of Linking and New Items in PIAAC 2023

Domain	Number of items		
	Trend	New	Total
Literacy	28	52	80
Numeracy	32	48	80
APS	/	65	65
Reading Components – Sentences	/	36	36
Reading Components – Passages	/	44	44
Numeracy Components – How many?	/	16	16
Numeracy Components – Biggest?	/	14	14

Note. Adapted from Ali and Robin (2025, p. 28). APS = adaptive problem solving.

Development of New Items

For PIAAC Cycle 2, new test items were created to reflect the updated frameworks and to expand the item pool, item types, and the range of item difficulties. Item development was guided by the framework specifications in order to achieve adequate construct coverage and varying levels of difficulty. Item developers used real-world content, such as articles, advertisements, and websites, and took advantage of the possibilities provided by the new tablet-based functionalities, such as drag-and-drop options. Item development includes defining scoring rules, that is, specifying which responses are coded as correct or incorrect. Automatic scoring was then implemented automatically in the international PIAAC software; this was a prerequisite for the adaptive algorithm applied in the cognitive assessment. Collaboration with expert groups and country reviews ensured that item development was aligned with frameworks and that the new items were culturally and linguistically appropriate.

tically appropriate for an international assessment of adults. In order to maintain trends with future cycles of PIAAC, care was also taken to avoid content that could easily become outdated. A large pool of items was developed. After analyzing the field trial results, the final item pool was selected for the main study based on psychometric properties and construct coverage.

Selection of Trend Items

Trend items from Cycle 1 were selected according to several criteria, including item quality, the distribution of items according to the dimensions specified in the updated frameworks, and the distribution across difficulty levels. Thus, items that did not have good item parameters in Cycle 1 or showed differential item functioning or problems with translations were not considered for inclusion in Cycle 2. Some items were excluded because they seemed too dated. The classification of the items was adapted to reflect the Cycle 2 frameworks, and it was ensured that all difficulty levels from Cycle 1 were reflected in the item selection.

After all items were selected and approved by the expert groups and the Board of Participating Countries (BPC), they were imported into the PIAAC Cycle 2 software platform and adapted as required. Adaptations included updates to user instructions to reflect tablet functionalities or adding a keypad for numeric entries. In addition, National Centres could correct errors and apply for national updates to the trend items during translation (see Section 2.4).

Item Layout and Navigation

The Cycle 2 item layout was adapted from Cycle 1 (cp. Lennon & Halderman, 2025). It maintained a consistent layout across cognitive domains but adapted elements from the Cycle 1 layout to match features familiar to present-day respondents. As in the first cycle, a vertical split-screen was implemented, with stimulus material located on the right and task or question information located on the left, allowing for better readability (see examples of items in English in OECD, 2024b and in German in Rammstedt et al., 2024b). The layout was developed to function effectively across different languages, including longer-text languages like German. Scrolling, for example, can affect item difficulty. It is therefore important for scrolling requirements to be as similar as possible across languages. In addition, the choice of layout considered the trend toward widescreen displays, ensuring it would remain viable for future PIAAC cycles. The location of the arrow and help buttons were changed to reflect the location of navigation icons on contemporary websites and widely used word processing programs.

Items were organized into topic-based units across all domains, each with one or more stimuli. In literacy and numeracy, respondents could navigate within a unit but not between units. In contrast, adaptive problem-solving units did not allow backward navigation within a unit due to the dynamically evolving nature of the tasks. These particularities were explained in domain-specific introductions. Additionally, some literacy and adaptive problem-solving units required navigating across multiple stimuli using tabs or website pages.

Item Response Modes

Item response modes were revised for Cycle 2 to account for the new tablet-based assessment (for details see Lennon & Halderman, 2025). Existing response modes (selecting information, highlighting text, and entering numbers) were assessed and adapted and a new response mode (“drag-and-drop”) suited to touch-screens was introduced. The response formats were familiar to respondents from using a smartphone or tablet. No specific ICT skills were necessary (Lennon & Halderman, 2025).

In Cycle 2, several specific response modes were implemented:

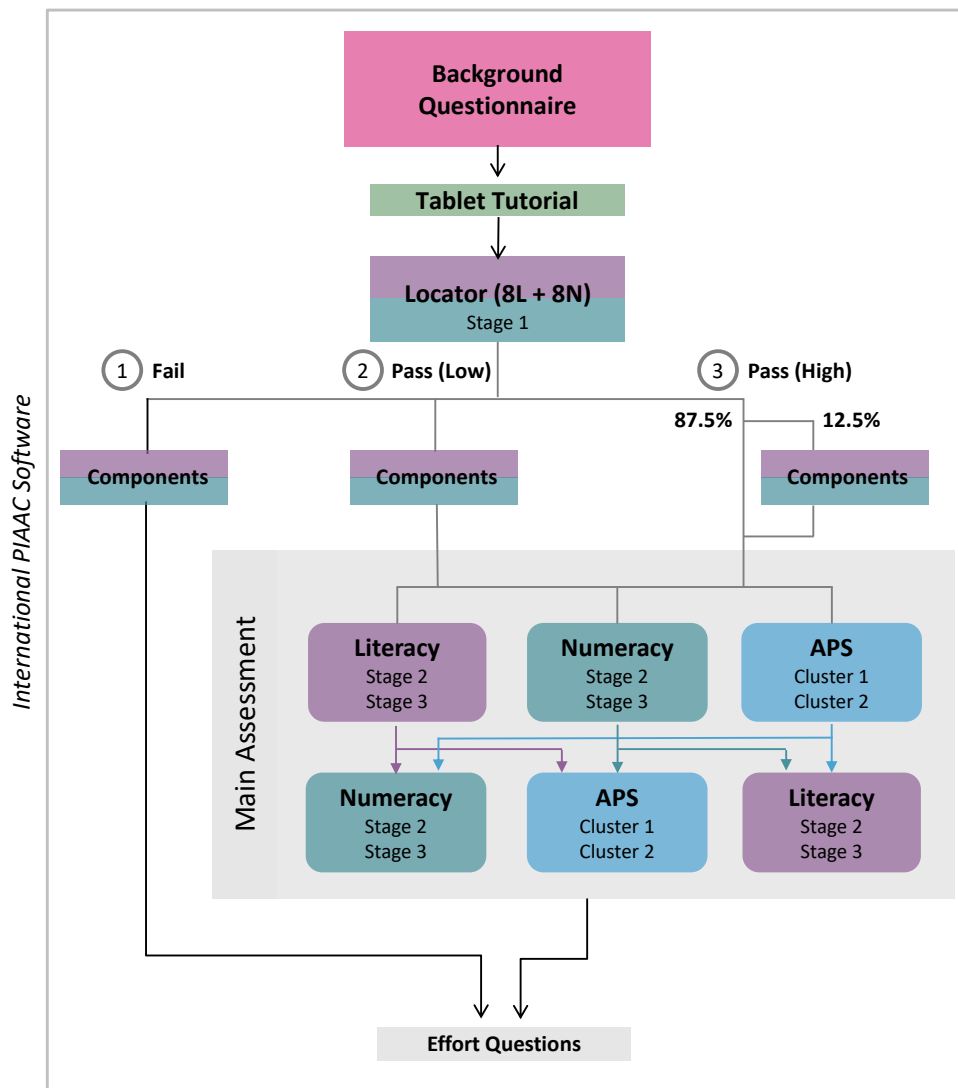
- *Tapping*, which allowed respondents to select text, graphics, or options on the screen.
- *Highlighting* was used in Cycle 1 literacy tasks and adapted to Cycle 2. When highlighting, respondents could mark words or sentences with a stylus pen. While highlighting items were necessary to link to Cycle 1 and earlier paper-based assessments, implementing and testing highlighting items across languages was very time intensive. In Cycle 2, highlighting was implemented for trend items but was replaced for newly developed items with different options, such as tapping on sentences.
- *Drag-and-drop* was used for items that required respondents to place choices in a specific order or move objects.
- *Numeric entry* was used when respondents needed to enter numbers as a response. Because the Cycle 2 assessment was tablet-based, respondents could not use a keyboard to enter numbers as they had done in Cycle 1. Instead, respondents were provided with an on-screen keypad for entering numbers.

Automatic Collection of Response Metrics

The international PIAAC software in Cycle 2 collected various response metrics. These included information about the response type and the results of the automatic scoring procedures coding responses as correct or incorrect. Automated response metrics also logged information about the time taken to respond to an item and the number of actions respondents carried out on a specific item. Furthermore, additional behavioural data was stored in log files for further analysis.

2.2.3 Assessment Design

The direct assessment contained several components (see Figure 2.2): (a) the tablet tutorial, (b) the locator test, (c) the components, and (d) the main assessment (which consisted of two of the three assessment domains, as described below). Two effort questions followed the assessment.

Figure 2.2: Assessment Design

Note. Adapted from Ali and Robin (2025, p. 23). For details on assessment stages and clusters see Section *Main Assessment*. L = literacy; N = numeracy; APS = adaptive problem solving.

Tablet Tutorial

The first part of the direct assessment consisted of an introduction to the tablet features and the assessment – the tablet tutorial. An extensive description of tablet actions, question and response types, and the layout of the assessment items provided all the necessary information to navigate the direct assessment. The tutorial explained which types of actions would be necessary to complete the assessment, such as tapping, dragging and dropping, and highlighting text. It also offered the opportunity to practice these actions prior to the direct assessment. Short animation videos showed how to perform all necessary actions and respondents could choose to watch the animations as often as they deemed necessary. Next, a general introduction showed the various question types found in the assessment (e.g., tapping on response option, tapping on sentence, or entering response into key-pad). Short animations showed how to enter responses to each question type. Fur-

thermore, the tutorial provided a description of the assessment item layout, showing where the directions, questions, and information required to respond to the question were located on the screen. The description also detailed how to navigate between items and introduced the help function and the online calculator (available for the numeracy assessment).

Locator Test

Following the tablet tutorial, respondents completed the locator test, the first stage (Stage 1) of the adaptive assessment design. The locator test comprised sixteen items: eight literacy and eight numeracy items. The literacy items included three Below Level 1 and three Level 1 items, as well as one Level 2 and one Level 3 item.¹¹ The numeracy items included two Below Level 1 and four Level 1 items, as well as one Level 2 and one Level 3 item (Lennon & Halderman, 2025). The eight items from each domain were completed consecutively and the sequence of domains was randomly allocated. Depending on the number of correct responses to the locator items, respondents were routed to the main assessment in one of three different paths (Ali & Robin, 2025; for path allocation, see Figure 2.3):

- Respondents routed to Path 1 only completed the reading and numeracy components.
- Respondents routed to Path 2 completed an assessment of reading and numeracy components, followed by a randomly selected combination of two skill domains.
- Respondents routed to Path 3 proceeded to the main assessment, where they were randomly assigned two of the skill domains. A randomly selected subset of the respondents in Path 3 (12.5%)¹² received an assessment of reading and numeracy components before completing items from the two skill domains.

Figure 2.3: Path Decision Rules

		Literacy locator scores		
		0-2	3-6	7-8
Numeracy locator scores	0-2	Path 1	Path 2	Path 3
	3-5	Path 2	Path 2	Path 3
	6-8	Path 3	Path 3	Path 3

Note. Adapted from Ali and Robin (2025, p. 30).

11 PIAAC literacy and numeracy levels range from Below Level 1 to Level 5. Higher levels correspond to higher literacy and numeracy skills.

12 In the original design, 25% of Path 3 respondents received reading and numeracy components. After the design was adjusted via a software patch to reduce the length of the assessment during fieldwork, the percentage was reduced to 12.5% (see Section 2.2.4).

Reading and Numeracy Components

The assessment design specified that all Path 1 and Path 2 respondents were assigned reading and numeracy components. In addition, to make sure that enough respondents with different proficiency levels completed the components, a randomly selected group of Path 3 respondents was allocated reading and numeracy components before continuing to the main assessment. The reading components assessment consisted of two sentence clusters and one passage. The numeracy component assessment consisted of two ‘how many’ clusters and two ‘which is biggest’ clusters. All respondents allocated to the components received both reading and numeracy components – the order in which they were assigned was randomly allocated.

Main Assessment

The following summary of the main assessment design is based on the detailed description in Ali and Robin (2025). In the main assessment, respondents were administered items from two of the three PIAAC skills domains: literacy, numeracy, and adaptive problem solving. Respondents in Path 2 and Path 3 of the main assessment completed two assessment blocks. In the first block, one of the three domains was randomly allocated to the respondents. The probability of being assigned to each domain was 33.3 %. In the second block, the second domain was randomly selected from the two remaining domains, each with a probability of 50 %. Thus, respondents completed tasks either from (a) literacy and numeracy, or (b) literacy and adaptive problem solving, or (c) numeracy and adaptive problem solving. The order of the domains was random. Each block was preceded by a domain-specific orientation describing specific response and stimulus options, such as highlighting for literacy, using the calculator for numeracy, or switching between tabs for adaptive problem solving.

The main assessment of literacy and numeracy used a hybrid design with an adaptive and a linear component. An adaptive algorithm was used to select the assessment items for 75 % of respondents. To ensure that a sufficient number of respondents with varying proficiencies completed each item, a linear assessment path was implemented for a randomly selected 25 % of respondents. The main assessment of the literacy and numeracy domains was divided into two stages, Stage 2 and Stage 3. Items were grouped into sets of items (testlets), whereby items were allocated to multiple testlets. The adaptive path of the literacy and numeracy assessment consisted of six possible testlets at each of the two stages. For Stage 2, there were three low-difficulty and three high-difficulty testlets. For Stage 3, there were two low-difficulty, two medium-difficulty, and two high-difficulty testlets. Based on responses to the locator items, the adaptive algorithm assigned one of the six possible Stage 2 testlets. Subsequently, based on responses at Stage 2, the adaptive algorithm selected a Stage 3 testlet that fit the difficulty level predicted for the individual respondent. In the linear assessment path, the Stage 2 and Stage 3 testlets were randomly selected, independent of performance.

Adaptive problem-solving items were allocated using a linear design only, since the locator did not provide enough information on expected proficiency in adaptive problem solving to realize an adaptive design. Respondents were assigned two randomly selected clusters (item groups) from a total of five possible clusters.

Items were grouped into clusters to ensure that the assessments of adaptive problem solving were of comparable difficulty.

Effort Questions

Two follow-up questions measuring subjective performance and effort were posed immediately upon completion of the direct assessment. The respondents were asked to rate the effort they put into working on the assessment items and how well they thought they had performed.

2.2.4 Administering the Assessment

After completion of the background questionnaire, respondents continued to the direct assessment. For the assessment, interviewers converted the tablet computer to a touchscreen-only tablet. The keyboard was disabled upon conversion. The interviewers then handed the tablet to the respondents, who used the touchscreen function (with a stylus or their fingers) for the self-administered direct assessment. Interviewers tried to set up a good working environment for the assessment and to limit external disturbances as far as possible. They were trained to be sensitive to respondents' concerns, which can easily arise in a "test" situation. Interviewers provided respondents with the necessary materials: a stylus, paper and pencil for notes, and a handheld calculator (for the numeracy domain, as an alternative to the online calculator integrated in the international PIAAC software). The interviewers were under strict instructions *not* to help with the assessment; they also had to ensure that no-one else helped in any way and that no other material or external sources were consulted. Translation support was *not* allowed (contrary to the background questionnaire). Interviewers were there to assist with the tutorial (if necessary) and with technical issues. Interviewers unobtrusively monitored respondents' progress, and, importantly, motivated respondents to persevere until the end of the assessment. Although short pauses were allowed, it was not permitted to pause the assessment and continue on another day; if for some reason the assessment had to be discontinued, this was a definitive breakoff.

The assessment was designed as an untimed test – the respondents could take as much time as they wanted to complete it. On average, the direct assessment in Germany lasted 75 minutes, which was longer than the international average of 60 minutes.

Adjustment of Direct Assessment During Fieldwork

In the early phase of fieldwork, analyses of the average duration of the assessment showed that it took significantly longer than expected (in Germany, around 83 minutes instead of one hour). In addition, interviewers reported that the long assessment had a detrimental impact on the fieldwork results. A quick reaction was required, and a software patch was developed by the international Consortium that reduced the length of the assessment to approximately 71 minutes in Germany. Most countries implemented the patch after two months of fieldwork (in Germany, the patch was deployed at the end of October 2022). The patch adjusted the assessment design as follows: (a) the share of Path 3 respondents who were randomly allocated reading and numeracy components was reduced from 25 % in the original design to 12.5 % in the updated design, (b) the number of units presented to the respondents at Stage 3 was reduced by one unit for literacy and two units for numeracy, and (c) for adaptive problem solving, the last unit in Cluster 1 and the first unit in Cluster 2 were removed.

In Germany, 26 % ($n = 1\,227$) of the PIAAC interviews were completed before and 74 % ($n = 3\,566$) after the update of the international PIAAC software.¹³

In Germany, 1.1 % ($n = 51$) of the respondents were routed to Path 1, 7.7 % ($n = 358$) to Path 2 and 89.7 % ($n = 4\,193$) of the respondents were routed to Path 3.¹⁴ The percentages were roughly equal before and after the update of the international PIAAC software.

2.3 Testing the Instruments

It was crucial to rigorously test the instruments prior to fieldwork and to ensure that they and the technical environment worked as intended. Each country received their own specific version of the PIAAC software, in their language and with their national adaptations. Although the PIAAC software was first tested at the international level, testing the national version was a core national responsibility. Testing was carried out in several iterations prior to the field trial, with national teams reporting bugs and errors, and the international Consortium re-issuing updated versions of the software. Specialized and end-to-end testing was carried out for the background questionnaire and the assessment separately. Several further rounds of testing were necessary for the main study, since changes were made to certain items and the domain introductions. The instruments were tested in three different environments: (1) on a testing portal, (2) in the international PIAAC software, and (3) in the final software constellation, after the international PIAAC software had been integrated into the Case Management System (this is a management software used during fieldwork, see Section 3.4).

2.3.1 Testing the Background Questionnaire

To ensure the correct functioning of the German background questionnaire, the National Centre tested the background questionnaire extensively, with a focus on the following key aspects: (a) correct implementation of changes between Cycle 1 and Cycle 2, (b) routing, (c) layout, (d) computation of derived variables,

¹³ Including doorstep interviews.

¹⁴ Excluding doorstep interviews; 1.6 % ($n = 73$) were not assigned a path (breakoffs).

(e) export files, (f) CAPI functionality and navigation, (g) stability of the CAPI system, (h) review of the final translation, and (i) overall flow of the background questionnaire.

The questionnaire was programmed in the so-called Questionnaire Authoring Tool. This software produced track-change documents which showed the changes implemented in every iteration. These changes were compared to the documented changes in the Background Questionnaire Adaptation Spreadsheet, an Excel file used as a documentation tool between the National Centre and the international Consortium. The correct implementation of all changes to the wording, routing, as well as additions or deletions of questions was then checked in the software.

Routing testing was carried out for each question in the background questionnaire individually. Every possible response was entered, and the subsequent routing was checked against the specifications. Due to the complexity of the routing, numerous paths through the background questionnaire had to be tested. These paths were then compared with a national BQ¹⁵ HTML file where the routings were implemented as hyperlinks, and with a graphical representation of the background questionnaire flow.

The layout checks involved ensuring that the screens appeared correctly, conformed with the intended layout, and that the text was accurate, complete, free of errors, and with correct font and color (e.g., black for text to be read out to the respondent and blue for the interviewer instructions).

All derived variables (these are variables that are automatically created by the CAPI system by combining information from different variables) were systematically verified. This was particularly important in Germany, given the number and complexity of national derived variables for education questions.

After the test runs, the corresponding export files were checked. This entailed ensuring that the correct data files were exported from the international PIAAC software onto the tablet computer, that all files could be opened, and that information stored in the files was accurate and complete.

Functionality tests included checking the navigation through the background questionnaire using the different options for navigation, that is, the keyboard or the touch functionalities (using a finger or a stylus). Help functions were also tested to ensure they were operational.

In order to test the stability of the CAPI software, testers tried to crash the system through several methods, such as testing very rapid user input (fast clicking or navigating back and forth very quickly).

Translation testing involved examining whether the wording used throughout the background questionnaire was harmonized (also through the different routings) or that each question was logical and understandable in the questionnaire flow. It also involved checking the so-called *dynamic text*. Dynamic text is used for questions that are worded differently depending on the responses to previous questions (e.g., past or present tense used depending on whether the questions refer to the last or the current job).

The overall flow was tested using predefined biographies. Internationally specified biographies were adapted for the German background questionnaire, and additional national scenarios were developed for national testing purposes. Mock

15 BQ = Background Questionnaire

interviews were also conducted. These tests verified that the correct questions were administered and that the content and numbering of showcards were consistent.

2.3.2 Testing the Assessment

The international Consortium provided testing scenarios to test the assessment instrument in the three different testing scenarios mentioned above (testing portal, international PIAAC software, final integrated national software constellation). These scenarios formed the backbone of testing. In Germany, additional testing plans were developed to ensure that all aspects of the assessment instrument were thoroughly tested. Testing protocols encompassed all key aspects of the assessment instrument:

1. *Content checks.* All items were reviewed in an item preview and later in the final international software. Testers checked for completeness, correctness, typos, grammar and spelling errors, adherence to the source in terms of font, layout, and readability.
2. *Functionality checks.* Various response modalities were available in the items across all three domains, such as selecting a single response option or multiple-choice options, drag-and-drop to order responses, tapping to highlight, or typing in numerical responses. Extensive testing ensured that all functions and navigation functionalities (forward and backward movement across and within units, navigation across multiple browser tabs) worked properly in all items.
3. *Scoring testing.* Extensive scoring testing was conducted in the online testing portal for all assessment items. To ensure that the scoring definitions were implemented correctly, each item was tested systematically by entering different correct or incorrect responses. For example, for numeracy items, numbers with decimal places were tested with different decimal separators and a varying number of decimal places to ensure that all possible options were scored correctly. Scoring testing was reconducted in the national version of the PIAAC software. In this process, the National Centre additionally checked exported result files to ensure that all entered responses were correctly scored and documented.
4. *Routing testing.* The National Centre systematically tested the international scenarios with predefined test cases for the adaptive design to determine whether routing was implemented correctly. The test cases specified the number of correct/incorrect responses that were required in order to be routed to a specific path.
5. *Stability checks of international PIAAC software.* The stability of the international PIAAC software was thoroughly tested throughout the assessment, for example by extensive tapping on the stimulus, tapping outside of the provided margins, attempting to complete the assessment extremely quickly, and letting the battery run out on purpose.

2.4 Instrument Translation

The objective of the PIAAC translation approach is to produce national instruments that are equivalent to the source while at the same time demonstrating a natural fluency in the target language. High quality instrument translation and adaptation are key to comparability. As such, the translation phase was a special focus both for the international Consortium and the National Centre. This section describes how translation was conducted in Germany. Some of the information on the standards and procedures pertaining to the translation, adaptation, and verification of the PIAAC survey instruments at the international level are based on Chapter 7 of the International Technical Report (Wayrynen et al., 2025).

The international Consortium established elaborate processes for linguistic quality assurance and linguistic quality control. International standards were defined and comprehensive guidelines for all translation steps and activities produced (e.g., process, good translation practice, content issues), including item-specific translation notes. Separate guidelines were available for the translation of the background questionnaire and the translation of the cognitive assessment, reflecting the different translation challenges encountered. For both instruments, double translation with subsequent reconciliation was strongly recommended in the PIAAC standards and guidelines (OECD, 2022). A training kit, including presentation slides, and additional supporting material were also made available to National Centres. PIAAC aims to achieve comparable measurements not only across countries, but also across cycles. Maintaining trend with the first cycle of PIAAC was therefore also a priority in this process. Different translation and verification approaches were developed for linking or trend items versus newly developed material.

The international Consortium created individual translation timelines for each country and provided translation tools customized for PIAAC.¹⁶ The overall approach included the following steps:

- translation training (international Consortium and National Centre)
- double translation (National Centre)
- team reconciliation (National Centre)
- linguistic verification (international Consortium)
- post-verification review (National Centre)
- layout and functionality checks, final review and final check (international Consortium and National Centre)

Translation was based on a U.S. English source and took place in 2019 in preparation for the field trial. After the field trial in 2021, the translations were reviewed, also taking psychometric results and flagged items into account. At this point only minimal changes were allowed.

It was the responsibility of the National Centres to translate and adapt the source instruments in accordance with the PIAAC best-practice standards. In Germany, the survey instruments were translated exclusively into German (as the official language). The German National Centre assembled a team of experts with a strong background in questionnaire and assessment translation for the team rec-

¹⁶ Translations were conducted using the OmegaT translation tool (with PIAAC Customizer).

conciliation of the two independent translations. Multi-day reconciliation meetings were held separately for the background questionnaire and each of the assessment domains.

One unique challenge of translating English instruments into German is the structural difference between the two languages: The German translation is generally significantly longer than the English source. The considerable length of the German language (compared to English and many other languages) can therefore impact on measurement (especially for the literacy domain), respondent motivation, and equivalence. The length of the translated text was therefore an important consideration, but it was crucial not to compromise faithfulness to the source version.

2.4.1 Collaboration with Other Countries

Sharing translations between countries translating into the same language was encouraged in PIAAC and there was a certain amount of collaboration and translation sharing between the German and the Austrian National Centres, and to a lesser extent with the Swiss National Centre. It was agreed between the German and the Austrian National Centres that each team would contribute one initial translation for the domains numeracy and adaptive problem solving and use the other Centre's versions as the second independent translation, thus resulting in cost savings on both sides. The initial translations of the background questionnaire and the literacy domain were also shared as an additional reference. However, due to each country having an individual translation timeline, the translation steps were not always aligned.

2.4.2 Translation Staff

The German National Centre contracted three translators.¹⁷ All were qualified professional translators with excellent command of English and German, translating into their mother tongue (German). Familiarity with translation tools was a prerequisite. The translators for the background questionnaire had many years of experience translating questionnaires. One of the translators for the cognitive assessment had previously translated test material for PISA and PIAAC Cycle 1; the second was new to assessment translation.

The staff at the National Centre was multilingual with excellent linguistic skills and experienced in the process of survey instrument translation. In addition, a national translation expert provided advice and support throughout the translation process.

¹⁷ One translator worked on both the background questionnaire and the cognitive assessment.

2.4.3 Translator Training

Attendance at the in-person international translation training was mandatory for national translator coordinators. The translator coordinators from the German team and the National Project Manager attended the international training. The internationally developed translation package was subsequently adapted for the national translation training and replicated for the three German translators. A two-day mandatory translation training was conducted in-person by the national translation team, covering all important aspects of the translation tasks and the PIAAC translation tools.

Translators were provided with extensive material, including documentation, forms, and software installation instructions. They were also asked to familiarize themselves with the survey material by first personally answering the questions (for the background questionnaire) or working through the assessment items (for the cognitive assessment) prior to starting their translation task.

2.4.4 Translation of the Background Questionnaire

Two different approaches are often followed to achieve comparability of questionnaire items across countries. One approach is to *ask-the-same-question*, also known as input harmonization. However, some demographic and socio-economic variables require *output harmonization*, whereby the measurement is obtained via country-specific items that are later coded into an internationally comparable target variable (e.g., Wolf et al., 2016). The majority of the items in the PIAAC background questionnaire conform to the ask-the-same-question approach. These are *translated* and *linguistically adapted* in accordance with PIAAC standards and best practices to obtain equivalence in the target language. For questionnaire items, translation also needs to maintain design features (this is especially relevant for response scales). This section focuses on the *translation* process for the background questionnaire. Some items, such as those measuring educational qualifications, need to be *structurally adapted*; this process is described in Section 2.1.2.

The international Consortium produced a source background questionnaire in U.S. English for PIAAC Cycle 2. This questionnaire was based on the Cycle 1 questionnaire and included some new modules and items as well as adjustments to some specific Cycle 1 items (see Section 2.1.4). Thus, a track changes document highlighting differences between the Cycle 1 and the Cycle 2 background questionnaire was also provided. The country-specific Background Questionnaire Adaptation Spreadsheet was used as a communication tool between the international Consortium and the National Centre. All translation and adaptations issues, discussions, and sign-offs were documented in this file.

For the translation of the background questionnaire, it is necessary to distinguish between the different types of items and the different translation approaches required by each type:

New items required double translation (by independent professional translators) and subsequent reconciliation. The PIAAC general translation guidelines (OECD, 2019a) strongly recommend a team or committee approach (see for exam-

ple Harkness, 2003) for the translation reconciliation. This approach was already followed in PIAAC 2012 in Germany and proved to be highly effective.

For the social and emotional skills module, however, a different process was in place (see Rammstedt et al., 2025, for a methodological description of the general translation procedure for this module). For Germany, a preexisting psychometrically validated BFI-2-S was available (Danner et al., 2019; Rammstedt et al., 2020) and was modified by the international Consortium for administration in PIAAC. National Centres could request changes to this version where they deemed that this was necessary; in Germany, only minimal changes were requested.

Trend items are questions taken from the Cycle 1 background questionnaires.

Strict trend items are identical to those in the Cycle 1 source questionnaire. For countries that participated in Cycle 1, such as Germany, a national review was carried out to identify potential errors or necessary updates, for example textual revisions resulting from language evolution or changed social reality. To ensure that countries only make strictly necessary (and not merely preferential) changes, countries had to justify their change requests and submit these to the international Consortium for approval. Actual changes could only be implemented by the international Consortium. Only very few changes were made for the German background questionnaire.

Soft trend items are Cycle 1 items that were changed to a certain extent by the international Consortium. These international changes had to be reflected in the national translations.

Adapted items denote structurally adapted questions. These were adapted in German and back translated to English for the communication and sign-off process with the international Consortium. Therefore, these items were excluded from the translation task.

National extensions are country-specific questions that can take up to five minutes of questionnaire time. These were developed in German and back translated to English for documentation purposes.

The translation procedure for the background questionnaire complied fully with the PIAAC standards and recommendations. Two professional translators produced independent translations using translation files and tools that were made available by the international Consortium. Translators were encouraged to provide translation comments (e.g., pointing out translation difficulties, specifying alternative options, or explaining reasons for deviations from standard wording). The two translations were reconciled using a team approach. The reconciliation meetings lasted several days and brought together the lead reconciler, translators, key members of the national team, and GESIS experts on translation, questionnaire design, education, and socio-demography. The two translations were reviewed minutely by the group and a final version was decided upon. After the reconciliation meeting, the National Centre conducted comprehensive final checks, notably ensuring consistent wording (e.g., key terms or phrases), looking out for routing specifications and target groups (e.g., particular care was required when dealing with dynamic text), and taking the CAPI administration mode into account (i.e., questions are read out by interviewers). Moreover, the National Centre rechecked all translations against the item-by-item guides before submitting the translations to the international Consortium for verification.

The verification of the background questionnaire translations carried out by the international Consortium included checks for functional equivalence of translation, linguistic correctness, consistency, and compliance with the PIAAC guidelines. For certain issues, the National Centre responded to the verifier comments, providing additional information or justifications for specific translations. The iterations between the National Centre and the international Consortium culminated in an agreement and final sign-off.

After verification, the German background questionnaire was previewed and tested in the actual interview environment (BQ preview tool). Checking the translations “in context” was an additional important aspect of the national translation quality control.

2.4.5 Translation and Adaptation of the Cognitive Assessment

For the cognitive assessment, the translation goal is to achieve “cognitive demand equivalence” (Wayrynen et al., 2025, p. 13). This translation task is different to that of the background questionnaire, since it requires awareness of the specific drivers of item difficulty. Translation entails, for example, mirroring the saliency of distractors in the translation, or ensuring that the same cognitive processes are triggered in the different language versions. At the same time, the national item versions must be culturally appropriate and the language fluent, natural, and at the same level of linguistic difficulty as the source.

The international Consortium produced general translation and adaptation guidelines for the cognitive instruments, laying out the translation standards, process, and materials for the translation of the cognitive items (OECD, 2019b). In addition, the international Consortium provided item-by-item guides, item information files (with the scoring definitions for each item, that is, the specification of correct responses, as well as information on estimated difficulty and response mode), checklists, and errata. The so-called Verification Follow-Up Form, a country-specific file containing information on each item, translation and adaptation issues, layout checks, and sign-offs, was used as a documentation and communication tool between the international Consortium and the National Centre.

Analogous to the background questionnaire, the international source assessment items were in U.S. English and different processes were in place for *new* versus *trend* items (cp. OECD, 2019b; Wayrynen et al., 2025).

New items required double translation by independent professional translators and subsequent reconciliation. For the *literacy* domain, two translators from Germany were contracted to produce double national translations, due to the importance of finding appropriate linguistic solutions for the German population. The initial Austrian translation was also available as an additional reference. For the domains *numeracy* and *adaptive problem solving*, one translation was produced by a translator from Germany and one by a translator from Austria (contracted by the Austrian National Centre). The Swiss German version was available as a supplementary reference for the adaptive problem-solving domain. All the items for the domain adaptive problem solving and the numeracy components were new items.

As the reading components items were administered on paper in Cycle 1 and transferred to the tablet-based mode for Cycle 2, the international Consortium adapted the material to the new mode where necessary with, for example, new instructions. The adapted material was translated into German and reviewed by the German National Centre. Furthermore, new reading component items from PISA were added to the reading component item pool. The German translations of the PISA items were also reviewed by the National Centre and changes were requested where necessary.

The tablet tutorial, the orientation, and the help modules were newly produced and needed to be translated. For this material, one translation was contracted by the German National Centre and one translation was provided by the Austrian National Centre. The translated material was carefully reviewed and finalized by the German National Centre. The translations underwent the same verification process as the translation of the cognitive instrument.

Trend items, that is, literacy and numeracy items from Cycle 1, were available in Germany. These items underwent some modifications to adjust them to the tablet administration mode, which was realized in a new technological environment. This especially impacted the instructions which were revised by the international Consortium. These international changes had to be incorporated into the translation. National Centres were also required to review all items to see if any error corrections or linguistic updates were required. Any proposed changes had to be justified and submitted to the international Consortium for approval. Only necessary, non-preferential changes were allowed. All trend items were carefully reviewed, proofread, the layout and scoring checked and, if necessary, adapted and documented.

The first step for all key translation and reconciliation personnel was to preview the item, read it through, and work through it themselves. The technical environment for PIAAC Cycle 2 allowed items to be previewed on a portal, which was a significant improvement to the translation process in Cycle 1. Items could be easily previewed in their different stages, and this allowed a much more thorough and holistic evaluation of translation alternatives. The second step was to review the scoring rules to understand what was regarded as correct and incorrect; this was helpful to identify distractors or traps and understand how the item works. Special care was taken to maintain equivalence with the source.

The translation procedure for the cognitive assessment complied with the PIAAC standards. The National Centre in Germany followed the team reconciliation approach for the cognitive instrument (Zabal & Behr, 2018). This method was also adopted in Cycle 1 and proved to be effective (Zabal et al., 2014). The meetings brought together the German translator/s, the national translation expert, key members of the National Centre, and the national domain expert (only for numeracy). During the reconciliation meeting, the two translations were reviewed sentence-by-sentence. The strengths and weaknesses of each translation were scrutinized, and new versions suggested. Adherence to the guidelines was closely monitored. For the literacy domain, a focus was on maintaining syntactical and vocabulary difficulty in the German translation (e.g., by comparing word frequencies in the source and the translation). Adaptations, translation challenges and the rationale for specific translation solutions were documented. Reconciliation work continued beyond the meetings. As a follow-up, the national team carefully reviewed

the reconciliation results, the translations (including flow and fluency), the overall consistency and harmonization (e.g., consistent directions), and compliance with item-by-item guides (e.g., whether identical word matches and synonymous word matches in the source were appropriately reflected in the translation). The review involved a careful layout check and final proofreading, after which the documentation was concluded. The fully bilingual lead reconciler performed the final check by working through each assessment item both in the source and in the German version to verify cognitive equivalence.

The verification of the cognitive assessment by the international Consortium included checks for functional equivalence of translation, linguistic correctness, consistency, and compliance with the PIAAC guidelines. For certain issues, the National Centre responded to the verifier comments, providing additional information or justifications for specific translations. The iterations between the international Consortium and the National Centre continued until an agreement was reached and the translations were signed off.

2.4.6 Review After Field Trial

In 2021, all countries carried out a field trial. One of the aims of the field trial was to test the functioning and equivalence of the survey instruments. Various analyses were carried out by the international Consortium to evaluate item functioning, timing, and routing. Items with poor psychometric properties or deviations between the national version and international results were flagged. Some modifications to the instruments were specified at the international level. For the background questionnaire, it was decided to drop several questions and to introduce some modifications reflecting changes due to the COVID-19 pandemic, such as including examples on remote work in several questions.¹⁸ Some changes were also introduced to the cognitive assessment, for example, the tutorials and orientations for the cognitive assessment were reworked. These international changes needed to be reflected in the national translations.

After the field trial, the National Centre carefully reviewed all the national instruments. The German and Austrian National Centres shared the international analysis results and reviewed flagged items and translations in a joint meeting. Some translations were slightly adapted following this review. Some smaller changes were made to the German background questionnaire for the main study. For example, the lists of most common languages and countries were updated to reflect the more current Microcensus data, the Microcensus 2020. For both the background questionnaire and the cognitive assessment, only a few wording revisions and some punctuation and typographical errors were identified. All changes in this phase were implemented by the international Consortium after they had signed off the national change requests.

¹⁸ For example, questions on the normal work environment and number of employees.

3

Preparation for Fieldwork

When preparations for the main study data collection began, it was clear that there would be many challenges to overcome. For one, given the nationally and internationally declining response rates (e.g., Beullens et al., 2018; Brick & Williams, 2013; Jabkowski & Cichocki, 2025), reaching the international PIAAC response rate targets of 70 % – or at least 50 % – seemed increasingly less realistic. To add to this, the effect of the COVID-19 pandemic was still ongoing, and it was unclear how this would impact on participation. In Germany, the General Data Protection Regulation (GDPR) had heightened sensitivity to issues of data privacy. Thus, privacy topics were becoming even more relevant; there were even instances of supervisory authorities limiting certain survey activities. And finally, the PIAAC interview itself, with an estimated administration length of two hours and with a direct skills assessment, added its own specific challenges. In a nutshell, a thoughtful and detailed fieldwork preparation was paramount. With this in mind, it was crucial to launch an effective and convincing outreach approach and also to recruit and train the best possible fieldwork staff (Sections 3.2 and 3.3). Other elements that also had to be specified and established prior to the start of the fieldwork were the general workflow and a tailored Case Management System to streamline this workflow and guide interviewers in their daily task (Section 3.4), as well as the hardware and technical implementation (Section 3.5). Before describing these components of fieldwork preparation, this chapter first summarizes another key pre-fieldwork activity, implementing the best possible sample design.

3.1 Sample Design and Selection

All countries participating in PIAAC implemented probability-based sample designs to obtain nationally representative samples of the PIAAC target population: adults aged 16 to 65 (OECD, 2022). Each country prepared detailed documentation of its sample design and selection procedures. The international Consortium carefully reviewed these materials and conducted a series of quality control checks throughout the sample selection process. Ultimately, all national sampling plans and activities were subject to the Consortium's approval (Van de Kerckhove, Li, et al., 2025).

PIAAC in Germany implemented a registry-based, two-stage, stratified and clustered sampling design. At the first stage, municipalities were selected as primary sampling units (PSUs), followed by the selection of persons at the second stage. This approach ensured that every person in the target population had a calculable and non-zero selection probability.

This section provides an overview of the target population, sampling frame, sample size, sample design, and selection procedures – both from an international and a national perspective. Information on all internationally defined standards and requirements for sampling is based on both the Technical Standards

and Guidelines (OECD, 2022) and the International Technical Report (Van de Kerckhove, Li, et al., 2025); these sources will not be referenced further.

3.1.1 Target Population, Sampling Frame, and Sample Size

The PIAAC *target population* comprises adults aged 16 to 65 living in private households with permanent residence in the country at the time of data collection – regardless of their citizenship, legal status, or language. Adults living in institutions, such as prisons, nursing homes, or military facilities, are excluded from this population. In countries using registry samples (such as Germany), target persons were considered eligible to participate in PIAAC if they were between 16 and 65 years old as of January 1, 2023 – the midpoint of the data collection period.

The *sampling frame* refers to the list or register from which a sample is selected. A high-quality sampling frame is essential to ensure adequate coverage of the target units. It must include accurate information on both the location and the measure of size of the target units. To meet the international PIAAC standards, the sampling frames were required to be of high quality and cover at least 95 % of the target population; each selection stage in a multi-stage sampling design required a separate sampling frame.

In Germany, sampling consisted of two selection stages, each with separate frames. For the first stage, the most recent version of the list of all municipalities in Germany was used – at the time of sampling, it was the version dated June 30, 2021, compiled by the German Federal Statistical Office.¹⁹ Approximately 0.06 % of the PIAAC target population was excluded, as municipalities located on islands in the North or Baltic Sea were inaccessible to interviewers.

The second-stage frame was based on the most recent register information available in 2022 from the population registries of the municipalities selected at the first stage. In Germany, all individuals are required to register their residence, and the frame is continuously updated through the official registration and deregistration of residents. However, not all individuals in the target population are necessarily registered in their local population registry. The estimated share of illegal immigrants – who are not officially registered – is around 0.5 %. In addition, the long data collection period of several months may have introduced potential issues with address validity. Some selected target persons had moved by the time the interviewer made the first contact, and for a portion of them, no new address could be located. This situation applied to approximately 3 % of the German sample after data collection was concluded. Nevertheless, Germany's overall non-coverage rate remained below the internationally defined threshold of 5 %.

Depending on the sample design, the PIAAC main study required a minimum *sample size* of 4 000 to 5 000 completed cases to meet the analysis goals.²⁰ In countries using a clustered two-stage design, such as Germany, a minimum of 4 500

19 The Federal Statistical Office provides access to the lists of municipalities in Germany (*Gemeindeverzeichnisse der Bundesrepublik Deutschland des Statistischen Bundesamtes*) for various reference dates on its website www.destatis.de.

20 For the definition of a completed case, see Section 4.5.

completed cases was necessary. This requirement was fulfilled: The final net sample in Germany comprised 4 731 completed cases.

3.1.2 Sample Design and Sample Selection in Germany

According to PIAAC standards (OECD, 2022), a self-weighting sample design was required to obtain a sample representative of its target population. Countries were expected to follow established scientific sampling principles to maximize precision while keeping data collection costs within reasonable limits. One of the most stringent requirements was that no substitution of sampling units was allowed at any stage of the sampling process.

PIAAC in Germany implemented a registry-based sample design, which is the methodologically best option (Häder, 2015). This approach was also used in PIAAC Germany 2012 and is common in other high-quality national social science surveys, such as the German General Social Survey (ALLBUS) or the European Social Survey in Germany (European Social Survey, 2019; Wasmer et al., 2017; Zabal et al., 2014). The field agency was responsible for carrying out all steps of the sample selection.

First Selection Stage: Municipalities

The first selection stage included a stratified random sample selection of municipalities with a probability proportionate to their size. To select the municipalities, the field agency stratified them based on characteristics such as federal state, administrative regions, and degree of urbanization. After the stratification, an algorithm subsequently allocated sample points to a stratification tableau. These sample points served as operational units for fieldwork planning. Each sample point represented a fixed number of target persons to be selected from the registry of the chosen municipality at the second selection stage.

The measure of size (MOS) for each municipality was based on the total number of persons in the target population. The allocation of sample points to the stratification cells occurred in proportion to the MOS of the strata. The MOS for each stratum was the sum of the MOS of all municipalities within that stratum. To allocate sample points to the strata, Cox's controlled rounding algorithm was applied (Cox, 1987). This procedure ensured that the distribution of selected municipalities mirrored the overall population distribution across federal states, administrative regions, and degrees of urbanization.

Once sample points had been allocated to strata, the algorithm randomly selected one municipality per sample point. However, municipalities with a large population could be chosen more than once in strata with more than one allocated sample point. Thus, these municipalities were represented by more than one sample point. For PIAAC Germany 2023, 320 sample points from 275 municipalities were selected using this sampling approach.

Second Selection Stage: Persons

After completing the first stage of sample selection, the field agency contacted the authorized registry office staff in the selected municipalities. The registry office staff was informed about the PIAAC survey and the request for sample selection. Along with a letter, the mailing included the following documents: a letter from the National Centre, a letter from the Federal Ministry of Education and Research, a study factsheet, instructions for the sample selection process, and preliminary versions of the data protection information sheet and the study flyer for target persons. The registry offices in the selected municipalities were instructed to (a) randomly select 68 persons per sample point from the target population and (b) provide personal information (e.g., name, address, age, gender, citizenship) for each selected person.

Past surveys in Germany repeatedly encountered problems with non-cooperative registry offices that were unwilling to carry out the sample selection at all. To address the various reasons for refusal given by registry offices (e.g., capacity constraints, prioritization of other administrative tasks), several measures were implemented in advance. For example, (a) ample time was allotted for providing the selected person addresses, which allowed for some flexibility in scheduling and (b) all municipalities received detailed information about PIAAC, together with a clear justification of its public interest. Where these steps proved insufficient, the sampling department at the field agency – and if necessary, the project directors at the field agency – maintained close communication with the registry office staff and addressed specific concerns (e.g., data protection issues) on a case-by-case basis. The National Centre supported this process and provided documents and input as required.

Some municipalities imposed specific constraints on the delivery of the sample, which were adhered to by the field agency. These constraints included, for example, the timely deletion of the contact data of refusers or limiting reissues of non-interview cases. Altogether, the comprehensive targeted efforts and supportive measures by the field agency and the National Centre were successful, and it was possible to comply with the strict requirement of avoiding any substitution of sampling units.

Over the course of several months, each of the selected municipalities ultimately delivered a sample of target persons from its local population, providing names, addresses, and additional personal information for these individuals. The field agency checked these samples for completeness and consistency (e.g., to identify duplicates or to ensure the correct target population was represented). In cases where some information was missing or incorrect, the field agency notified the registry offices requesting the transmission of corrected data.

Based on assumptions about the expected response rate for Germany, it was decided to assemble an initial gross sample of 11 200 persons and a reserve sample of 5 760 persons. After the cleaning process, the field agency selected a stratified sample of individuals from the full pool of delivered cases, drawing 35 sample units per sample point for the initial gross sample (35 cases x 320 sample points = 11 200). First, for each sample point separately, all (cleaned) cases were allocated to a 12-cell matrix (six age groups by gender). Subsequently, an optimization algorithm selected 35 persons per sample point, ensuring that the socio-demographic structure of the federal state in which the municipality was located

was well represented. The same approach was replicated for the reserve sample, with 18 persons per sample point.

3.2 Outreach

One of the challenges of survey implementation is developing a good strategy to address the selected target persons in order to gain their cooperation. For PIAAC Germany 2023, a comprehensive outreach strategy was established with the objective of achieving the highest possible response rate and meeting the fieldwork target of completed cases. An assortment of outreach materials and an incentive strategy were developed to stimulate survey response and reduce nonresponse.

3.2.1 Information Material for Target Persons

To capture the interest of the very heterogeneous target population, engaging and varied information material was produced. A lot of thought and care went into the development of the different information resources. The intention was to provide all relevant information and to cater for different tastes, so that each target person would find something in the mix that would appeal to them.

Advance Letter

The advance letter, that is, the letter which is sent per postal mail to the target person inviting them to participate in the survey, is the first point of contact with the target person. The first impression it makes is an important factor in whether or not the target person decides to participate. Prior to the field trial, the National Centre commissioned a cognitive pretest to evaluate three different advance letters designed specifically for PIAAC in Germany. Key findings of the pretest included (cp. Martin et al., 2022): (a) short letters are better, (b) a structured format with subheadings helps recipients to process the content of the letter, and (c) including the logo of the sponsor (the Federal Ministry) in the letter adds to its legitimacy and emphasizes the importance of the study. The results of the pretest informed the final design of the advance letter.

The advance letter included the following key elements (see Appendix B for the German original and the English back translation):

- personalized letter
- information on PIAAC, sponsor, and responsible institutions
- description of target person selection
- importance of participation and voluntary nature
- data privacy, reference to data protection information sheet
- incentives
- name of interviewer
- hotline number of field agency

To limit the letter to one-page, further information was given in a study flyer included in the invitation package sent to target persons (see Section 3.2.3). A few

communities had special requests that needed to be catered for in the advance letter (e.g., that address data had been transmitted in accordance with the Federal Act on Registration). For target persons who were under 18 years of age, an information package was also sent in parallel to the parents of the target person explaining that their child had been invited to participate in PIAAC.

Study Flyer

A study flyer was professionally produced and included in the invitation package sent to target persons via postal mail. The flyer introduced PIAAC as an international OECD study, named all institutions involved in funding and implementing the study, and highlighted the relevance of the study and participation. The flyer also briefly described how target persons were selected and emphasized that participation was voluntary. It gave a short account of the interview process, the approximate length of the interview, and illustrated tasks that respondents might be asked to complete. Information about the incentive was prominently placed. The flyer contained information on data privacy and provided the hotline number of the field agency that respondents could call for further information as well as the link and a QR code to the respondent website at GESIS.

Data Protection Information Sheet

In accordance with data protection regulations (e.g., GDPR), the data protection information sheet provided target persons with clear and concise information on how their personal data would be collected, processed, and protected. The following information was provided: (a) objectives and implementation of PIAAC, (b) sponsors and relevant organizations, (c) how individuals were selected for PIAAC, (d) voluntary nature of participation, (e) processing, storage, and use of personal data, and (f) information on rights and how to exercise these. The data protection information sheet was sent with the advance letter and was also handed to the respondent when obtaining consent prior to the interview.

Study Brochure

Interviewers were equipped with additional at-the-door instruments they could use flexibly to encourage participation. The main at-the-door instrument was a professionally designed study brochure that contained more extensive information about the PIAAC interview than the flyer did. The brochure included a preface by the National Project Manager, answers to frequently asked questions such as when the interviews would take place, how target persons were selected, and whether prior knowledge was necessary for participation. The brochure also included short descriptions of example assessment tasks, and information about the interview process, data privacy, and protective measures taken with respect to COVID-19.

Endorsement Letter

An endorsement letter was provided by the Federal Ministry of Education and Research that could be handed over at the door. This was especially appropriate for target persons who questioned the study's legitimacy.

Translations

For target persons with language barriers, translations of the advance letter, the data protection information sheet, and FAQ were made available in Arabic, English, Polish, Romanian, Russian, and Turkish. These were usually used flexibly by interviewers at the door and were also accessible via the respondent website.

PIAAC Video

Another at-the-door instrument was a professionally produced short video about PIAAC. This video was installed on all PIAAC tablets so that the interviewers could show it to respondents. It was also embedded in the PIAAC respondent website (see below) and uploaded to the GESIS YouTube channel, ensuring that respondents could find it themselves. The video was viewed over 500 times on YouTube. Several interviewers also showed the PIAAC video to respondents on their mobile phones.

PIAAC Respondent Websites

In addition to the information package sent to respondents and the at-the-door instruments provided to the interviewers, two PIAAC websites, specifically tailored to the respondents, were created for the main study. The PIAAC website at GESIS provided responses to frequently asked questions, such as the selection process or what happens during the interview. Furthermore, respondents could find information about the interview process and a wide selection of example tasks. The PIAAC video and respondent hotline information were also prominently posted on the site. The field agency's PIAAC website provided information on the survey as well as contact information for the PIAAC team at the field agency.

Press Release

A press statement, announcing the start of the PIAAC data collection, was sent to local newspapers nationwide. Prior to the main study, a list was compiled with email addresses of over 500 local news media from each municipality that was sampled for PIAAC. During fieldwork, the press release was emailed to local media in the municipalities selected for roll-out, ensuring that respondents would receive advance letters at the same time as reading about the study in the local news. However, only few news media published information about the start of the PIAAC interviews in their municipality.

3.2.2 Incentives

Respondent incentives are a well-established measure to increase response rates. In Cycle 1 of PIAAC, a 50-euro conditional cash incentive was offered in Germany. This decision was partly based on an incentive experiment carried out in the Cycle 1 field trial, which compared three different monetary incentives. The highest incentive, 50 euros, resulted in the highest response rate (Martin et al., 2014). Given the high respondent burden of the PIAAC interview, it was also considered a fair token of appreciation. Since response rates further decreased since the

Cycle 1 data collection in 2011/2012, and the impact of the COVID-19 pandemic on survey participation was unclear, the question for Cycle 2 was whether a conditional incentive of 50 euros would be sufficient. Blohm and Koch (2021) showed that for the German General Social Survey prepaid monetary incentives were more effective in increasing response rates than promised monetary incentives. They also found no systematic effects on sample composition. The National Centre in the Netherlands carried out an incentive experiment as part of their PIAAC Cycle 2 field trial (Buisman, 2021, personal communication): They compared two unconditional incentives (20 and 30 euros) with and without an additional 5-euro prepaid incentive. The highest response rates were achieved with the combination of a 5-euro prepaid and a 30-euro promised incentive. These results suggested that adding a 5-euro prepaid incentive to the 50 euros promised incentive could help boost response rates in PIAAC Germany. At the same time, there were practical and ethical issues to bear in mind, given that unconditional incentives are a cause for considerable controversy.

With a view to the challenging PIAAC response rate standards, the decision was made to offer 5 euros as an unconditional monetary incentive and 50 euros as a conditional incentive (both in cash). No conditional incentive was offered for the extremely short doorstep interview. As a reaction to difficulties encountered during fieldwork, the incentive was subsequently increased to 80 euros for the refusal conversion/reissue phase.

3.2.3 Invitation Package

The target persons were sent an invitation package per postal mail with a real postage stamp and three logos on the envelope: PIAAC, GESIS, and Kantar Public. The envelope contained the one-page advance letter, as well as the data protection information sheet and the PIAAC flyer, which were bundled together, with a 5-euro bill prominently placed on top. Inside the bundle was a sticky note block with the PIAAC logo on the front and the GESIS and Kantar Public logos on the back. The sticky note block added bulk to the invitation package and provided a tactile stimulus that made opening the envelope more interesting. The repeated use of the PIAAC logo was intended to make the study more familiar and trigger a recognition effect when the interviewers made contact. The invitation package was designed to look professional, official, and trustworthy and to set itself apart from advertising mail. At the same time, it was also designed to appeal to the target person and invite them to open the envelope and peruse the contents.

The invitation package was to be sent at predefined dates to coincide very closely with the interviewers' first contact attempts. Good synchronization maximizes the effect of the invitation package in obtaining respondent cooperation.

The invitation package for the refusal conversion/reissue phase was similar, but with a tailored approach and fewer enclosures (e.g., no prepaid incentive and no sticky note block).

3.3 Fieldwork Staff

The quality of interviewers' work is a central determinant of data quality. In PIAAC, interviewers must fulfill three quite different roles: (1) gaining respondent cooperation, (2) administering the background questionnaire, and (3) monitoring the cognitive assessment. The first role requires interviewers to be good contactors, professional communicators, comfortable with social interactions, able to improvise, and gifted persuaders. The second role requires adherence to standardized procedures, abiding by the PIAAC rules, and being exact in reading questions and recording responses. The third role requires the ability to step back and let the respondents work on their own, while at the same time being attentive (monitoring duties) and thoughtful, for example, by trying to create the best possible atmosphere for the respondent to complete the assessment tasks.

It was important to ensure that the field agency not only had an excellent interviewer staff but also the appropriate management staff and structures to coordinate and monitor the data collection. The management staff for PIAAC at the field agency fulfilled the PIAAC standards. There were two field directors, three field managers and six supervisors available for the PIAAC study.

3.3.1 Interviewer Selection and Characteristics

The decision on how many interviewers to recruit for PIAAC was influenced by various factors. While it was important to hire a sufficient number of interviewers in order to achieve the data collection target goals, it was also necessary to limit the number because all PIAAC interviewers needed to be equipped with their own newly purchased PIAAC tablet. At the same time, constraints were also imposed by the field agency's reduced staffing levels following the aftermath of the COVID-19 pandemic. Interviewer attrition is rare for such projects, so only very minimal over-hiring was anticipated. After weighing the constraints, it was decided to target approximately 100 PIAAC interviewers. The selection of the best available interviewers for PIAAC was carried out with great care and based on the following criteria:

- experience in conducting large-scale surveys in the social sciences, and specifically surveys based on registry samples
- excellent track record, reliable and high performing (response rates)
- proximity to PIAAC sample points
- availability for trainings and during fieldwork period (taking workload of concurring assignments into account)
- preferably experience with PIAAC Cycle 1, PIAAC-Longitudinal, or PIAAC Cycle 2 field trial
- tenure with the field agency

The goal of recruiting 100 interviewers for PIAAC was achieved. Table 3.1 shows the distributions of gender, age, years at field agency, and employment status. Most interviewers in Germany work as freelancers. However, the field agency had access to a small staff of permanently employed interviewers. As far as possible, permanent staff were assigned to the PIAAC project. On average, there were more male

interviewers, interviewers tended to be older, and they had often worked for more than 10 years for the field agency. Some interviewers were willing to travel for their assignments, sometimes spending weeks at a distant sample point. Half of the interviewers had experience with one or more of the previous PIAAC data collections (PIAAC Cycle 1, the national PIAAC-longitudinal study, PIAAC Cycle 2 field trial).

Table 3.1: PIAAC Interviewer Characteristics

Interviewer characteristics	Number	Percent
Gender		
Male	60	59.4
Female	41	40.6
Age		
Up to 49 years of age	7	6.9
50 to 59 years of age	19	18.8
60 to 69 years of age	40	39.6
70 years and over	35	34.7
Experience at field agency		
Up to 3 years	9	8.9
4 to 10 years	31	30.7
11 to 20 years	37	36.6
21 years or more	23	22.8
Not specified	1	1.0
Employment status		
Freelance, not available for travelling assignments	35	34.7
Freelance, available for travelling assignments	46	45.5
Permanently employed	20	19.8
Total interviewers	101	100.0

Note. Source: Data from Kantar Public (2024). Includes one interviewer who passed away before any PIAAC assignment could be completed.

Some interviewers worked part-time, others full-time. They usually worked on other surveys in parallel, which was an important consideration during selection and case assignment. Most sample points were worked by one interviewer only.

3.3.2 Interviewer Training

The interviewer tasks for PIAAC are numerous and varied. They often differ from national standard procedures and are at times novel and complex. To collect comparable data, it is essential that interviewers within and across countries carry out the PIAAC interview strictly according to the study protocols. Comprehensive training of interviewers is therefore one of the quality pillars of fieldwork preparation in PIAAC (OECD, 2022). The international Consortium developed an interviewer training program and trained key national training staff. For PIAAC in Germany, the National Project Manager and the field agency's Project Director, both lead trainers for the national interviewer trainings, attended the international training in-person; other national trainers participated online. Because all lead trainers attended the international training, no additional national train-the-trainer sessions were required. Five supervisors were trained at a 6-hour PIAAC-specific supervisor training session; the sixth supervisor attended an entire 5-day interviewer training. The field agency also held specific training sessions for the technical and the survey hotline staff.

The international training material was translated, adapted, and extended for the national interviewer training. For example, the following additional country-specific training materials were produced: (a) using the national Case Management System, (b) contacting and documentation requirements, including using national disposition codes, (c) national specifics of the background questionnaire, (d) PIAAC tablet and technical tips, (e) ensuring data privacy, (f) gaining respondent cooperation, (g) administering the doorstep interview, and (h) dealing with COVID-19-related issues.

In Germany, three 5-day in-person interviewer trainings were conducted at a conference hotel between mid-August and mid-September 2022. The trainings were led by a team of three lead trainers, four assistant trainers, technical support staff, and additional home office staff from the field agency. This was a joint event of the field agency and the National Centre. In all, 104 interviewers attended the interviewer training. Group sizes ranged from 33 to 36 interviewers. All interviewers had previously attended mandatory general interviewer training conducted by the field agency. The PIAAC interviewer training covered a wide range of topics in the course of the 28.5 hours in-person training plus 7.5 hours self study:

- *home study exercise* (3.5 hours self study): preparation for training
- *introduction to PIAAC and survey overview* (4.5 hours): background information on PIAAC; study design; sample design; fieldwork planning and timeline; data privacy and confidentiality; ethical responsibilities; study material and incentives for respondents; contacting rules; role of interviewer; material and support for interviewers; COVID-19 safety protocols
- *general interviewer techniques and basic CAPI training* (3 hours): interviewing techniques; standardized administration of questionnaire; probing techniques (with hands-on exercises); international PIAAC software for questionnaire: screen set-up, navigating in the software, shortcuts, warnings, error handling, help functionality; PIAAC protocols for the background questionnaire (e.g., what should be read out in verbatim, interviewer instructions, handling show-cards, response formats, use of *don't know* and *refused*)

- *administering the background questionnaire* (5.25 hours): introduction to the background questionnaire and overview of the different sections; using interpreters in case of language problems; reading out questions and entering responses; introduction to questionnaire questions (interactive sessions), dos and don'ts; focus on key aspects such as educational qualifications; exercises on obtaining detailed information on occupation and industry
- *administering the cognitive assessment* (3 hours): assessment design; the four assessment modules (tablet tutorial, locator, components, main assessment); interviewer role in the administration of the cognitive assessment, creating a good setting; tablet handover; providing the respondent with the necessary material (stylus, paper and pen, calculator); instructing the respondent; monitoring progress; reacting when required, dos and don'ts; practical walkthrough; familiarization of interviewers with some items
- *doorstep interview and curbside observations* (1.75 hours): administration of the doorstep interview; handling of doorstep material; interactions at the doorstep (role play); registering curbside observations (background, protocols, practical handling, possible criteria, examples and group discussion)
- *national Case Management System (nCMS) and disposition codes* (2 hours in-person, 1 hour self study): technical introduction to (hybrid) PIAAC tablet; launching international PIAAC software; PIAAC-specific features of the nCMS; communication with home office; data transfers; consent protocols; audio recordings; documentation of contacts and interactions with respondent; protocols for coding non-interviews and (interim) disposition codes (incl. exercises)
- *gaining respondent cooperation* (2 hours): reasons to participate in PIAAC; recognizing and addressing target persons' concerns (different scenarios and role plays)
- *quality control and administrative procedures* (1.5 hours): PIAAC quality measures and quality control (full transparency for interviewers); importance of detailed notes about non-interviews for validation evaluation; requirements and procedures for audio recordings
- *practice interview* (5.5 hours): conducting the PIAAC interview (both roles: interviewer and respondent)
- *post-training practice interview* (3 hours): conducting two practice interviews (whenever possible with audio recordings) with individuals from personal network

Interviewers were given their PIAAC tablets and PIAAC material at the training session. Interviewer training consisted partly of instructor-led modules with questions and answers, as well as many practical sessions with hands-on elements and exercise modules. The walkthroughs of the background questionnaire and the cognitive assessment helped interviewers familiarize themselves with the overall design, questionnaire questions, and some of the cognitive items. There was also an opportunity to go through and practice the protocols for pausing or breaking off the interview in the international PIAAC software. Interviewers generally respond well to interactive learning concepts, and the discussions of at-the-door scenarios and role plays in the module on gaining respondent cooperation were perceived as very fruitful and stimulating. The latter was an especially crucial module given the challenges of obtaining cooperation for PIAAC. The practice

interview was a key element of training. It simulates a real interview situation, allowing interviewers to familiarize themselves with the PIAAC interview, not only in their role as interviewers, but also from the perspective of the respondent. This offered valuable insights. The practice interviews were observed by the trainers. During the joint wrap-up session, interviewers were invited to share their experience, receive feedback from trainers and other participants, and extract important lessons learnt.

An important part of interviewer training is the evaluation of the interviewers. To this end, the trainer team met daily to review overall training progress and performance at the individual level. Supplemental one-on-one sessions were offered to interviewers who were identified as needing more practice, or who requested further support. The final evaluation of individual interviewers was based on their overall performance during training, with a focus on observations of practice interviews, since these reflected how well training content had been understood and integrated. In a few cases, a need for additional support and monitoring was identified and addressed by the field agency. One interviewer broke off the training, another interviewer did not perform adequately and was excluded from the PIAAC assignments, and a third interviewer was dismissed by the field agency shortly after the interviewer training. In the end, 101 trained PIAAC interviewers were available for fieldwork after training; 100 interviewers completed PIAAC assignments.²¹

The multi-day interviewer trainings also contributed to interviewer motivation and commitment to the PIAAC project. One of the additional benefits of such trainings is that they offer informal opportunities for interviewers to talk to their peers and the trainer team. This is important to interviewers, as such in-person exchanges are rare. In addition, joint events were organized and these contributed to the development of a team feeling.

The interviewer trainings were evaluated by the interviewers. Satisfaction with the trainings was high and most interviewers felt well-prepared for fieldwork. Various interviewers emphasized their appreciation of face-to-face trainings: They were motivationally important, facilitated knowledge acquisition, and offered opportunities for individual in-depth exchange.

After the interviewer training, interviewers were required to conduct two practice interviews with individuals from their personal network to consolidate the skills and knowledge gained during training. They were also asked to review their training materials and the interviewer manual before heading to the first PIAAC interview. Questions posed by training participants were documented throughout the three training sessions; an FAQ document was subsequently collated and distributed to all interviewers.

In a number of cases, specific retraining or refresher trainings were carried out by the field agency during fieldwork. Furthermore, the National Centre prepared a video tutorial to refresh the information on asking about and recording educational qualifications correctly. An additional complete interviewer training was not required due to low interviewer attrition.

21 One interviewer passed away early during fieldwork and before completing any PIAAC interview.

3.3.3 Interviewer Manual and Materials

The international Consortium produced a comprehensive interviewer manual that was to be used as a model to develop national material (OECD, 2022). This manual was translated and extensively reworked and adapted. The German PIAAC interviewer manual included the following sections:

- *introduction and overview*: information on PIAAC, international context, national organizational structures, timeline, PIAAC instruments, interviewer tasks, ethical principles, confidentiality
- *materials*: overview of the materials for the PIAAC data collection, interviewer checklist
- *contacting and doorstep interview*: contacting procedures and rules, recording contact attempts, assigning disposition codes, documentation, addressing target persons, problems making contact, handling problems with locating target persons, procedures for movers, doorstep interview, curbside observations
- *interview workflow*: verification questions, obtaining consent, administering the background questionnaire, using an interpreter for the questionnaire, assessment design, administering the assessment, guidelines and rules, dealing with special groups, handling special situations, wrapping up the interview, post-interview questions
- *quality control*: validation and audio recordings
- *administrative procedures*: contact with home office
- *appendices*: additional information and materials, incl. advance letter, flyer, doorstep cards, FAQ, data protection information sheet, confidentiality agreement

Interviewers received the manual when they attended the in-person training. They were required to read through the manual before starting on their PIAAC assignments. Interviewers were also provided with other comprehensive materials for their work. A concise interviewer booklet with practical information relevant for the administration of the interview was a key reference for the interviewers. It included hands-on instructions for different situations, for example, pausing or breaking off an interview, handling technical problems, procedure for audio recordings, and COVID-19 safety protocols. It also included a list of the materials needed for the PIAAC assignment:

For the contacting phase:

- information material for the target person (advance letter, endorsement letter, PIAAC flyer, brochure, FAQ, data protection information sheet; if available, also translations of these documents)
- interviewer ID
- contact cards
- address log

For the interview:

- PIAAC tablet and accessories (incl. charger, stylus, extra batteries, protective pad)
- data protection information sheet (and translated versions)
- show cards (booklet form) for the administration of the background questionnaire
- calculator, paper, and pens for the administration of the cognitive assessment
- 50 euros in cash (respondent incentive)
- doorstep interview cards (booklet form)
- masks, disinfectants
- interviewer manual
- interviewer booklet

3.3.4 Interviewer Remuneration and Incentives

Remuneration for a PIAAC interview was above average and thus more attractive than for other surveys, reflecting the national importance of PIAAC and the complexity and length of the interview administration. The payment scheme for *freelance interviewers* was composed of a basic piece rate plus a dynamic component based on interview length, with incremental add-ons for interviews that were longer than 90 minutes (increments were paid up to 180 minutes, after that a flat rate was paid). The dynamic component ensured that respondents were given as much time as they required for the assessment, without the interviewer pushing them to finish or cutting them short. The basic piece rate took into account the contacting and administrative tasks that form part of the interviewers' work. For a partial interview (background questionnaire completed but breakoff in the assessment), the piece rate was lower, thus discouraging shortcuts. A reduced piece rate was paid for doorstep interviews. All expenses were paid. Exceptional efforts were suitably compensated (e.g., special schemes for interviewers who travelled to distant sample points).

From the start, a bonus was offered for each sample point with a response rate higher than 50 % and for sample points with the highest response rates (top 1/3). The criteria for the bonus were adjusted during fieldwork to continue motivating interviewers. Due to the unexpected and inordinate length of the PIAAC interview at the beginning of fieldwork, the base rate was increased considerably and paid for all interviews. Increasing the payment was also a measure to incentivize interviewers to prioritize working on PIAAC in a period with many concurring assignments for other surveys. Due to the slow fieldwork progress and the difficulties encountered in achieving the data collection goals (see Section 4.3.3), more individual interviewer incentive agreements were introduced during the last months of fieldwork in order to increase the intensity with which cases were worked.

Permanently employed interviewers received a salary plus expenses. A supplementary payment was introduced after the first months of fieldwork to compensate them for the challenges and length of the PIAAC interview. In addition, they were further motivated through various bonuses, which were awarded to especially active and successful interviewers. These bonuses were offered at different time points during data collection.

3.4 PIAAC Workflow and Case Management System

Interviewers have to carry out a wide variety of tasks according to specific protocols and are generally guided and supported in their daily work by a survey-specific Case Management System (CMS). At the same time, the CMS is a key tool for managing the data collection process on a day-to-day basis and for effectively overseeing survey operations by supporting supervisors in their daily operational tasks. For PIAAC, countries were expected to use a CMS with a range of functionalities, including:

- assigning cases to interviewers
- documenting contacts and contact attempts
- recording appointments
- recording disposition codes
- launching the interview (international PIAAC software)
- managing breakoffs
- providing data for monitoring and reporting purposes
- saving, extracting, and transferring the collected interview data

In Cycle 2, the international Consortium developed an automated international CMS (iCMS). Countries could either use the iCMS or, alternatively, their own national CMS (nCMS), provided it offered at least the same core functionalities as the iCMS. The iCMS allowed only limited flexibility and did not cater for national customizations. In addition, there were data protection concerns, since respondent data was to be stored on servers hosted by the international Consortium (and possibly outside of Europe). The German field agency already had a reliable CMS in place, which had the additional advantage that interviewers were accustomed to working with it. It was decided to adapt this management platform to create an nCMS that complied with the specific national requirements while including all core functionalities that were internationally required. The field agency was responsible for the smooth and stable integration of all internationally required components into the nCMS. They were supported in this work by the National IT Coordinator. The integration included developing appropriate interfaces with the international PIAAC software and ensuring a secure and reliable data export process. Before the start of fieldwork, the nCMS underwent extensive testing – both as a stand-alone system and in combination with the integrated components – to ensure it would function as intended on the interviewer tablets.

The nCMS was a dedicated application installed on the interviewer PIAAC tablets, to be used by interviewers to perform administrative tasks before, during, and after each interview, as well as to document nonresponse cases. From the interviewer's perspective, the nCMS was designed as a structured, questionnaire-like interface composed of multiple screens that guided them through their workflow in a user-friendly and consistent way. The nCMS led the interviewers through the following key activities:

- documenting contact information
- assigning disposition codes
- entering information for follow-up attempts
- recording curbside observations

- selecting the *main interview*, the *doorstep interview*, or the *documentation of nonresponse cases*
- verifying the identity of the sample person
- obtaining informed consent to conduct the interview
- launching the international PIAAC software and, where applicable, the audio recording software
- managing breakoffs
- capturing respondent's telephone number for validation purposes (if possible)
- recording interviewer observations

The nCMS stored a record of every case assigned to an interviewer with all relevant information per case. For interviewers, the work on each case began with the contacting phase, that is, personally going to the address of a selected target person. On-site, interviewers were expected to attempt to make personal contact with the target person. In many cases, no contact could be made or only with another household member. Interviewers were required to document every contact attempt or actual contact in the nCMS (see Section 4.2.1 for information on contacting rules and contact documentation).

When visiting the target person's address for the first time, and before doing anything else, interviewers were required to complete six short curbside observation questions (see Section 4.6). The nCMS was programmed to ensure that the observations were recorded before any contact was attempted (i.e., ensuring that a consistent approach was followed for both respondents and nonrespondents).

Three distinct scenarios could result from the contact phase, each leading to a different workflow within the nCMS. Interviewers could select the appropriate path: conducting the *main interview*, conducting a *doorstep interview*, or *documenting a nonresponse case*.

Path Main Interview

To conduct the main interview, a series of follow-up screens were built into the nCMS, guiding interviewers through several preparatory steps before launching the international PIAAC software and beginning the interview. First, interviewers verified that the person they intended to interview was indeed the correct target person, by checking their name, gender, and age, and confirming this information in the nCMS. If the verification indicated that the person was not the target person, no interview was to be conducted, and the interviewer was required to identify the correct target person. If verification was successful, interviewers were prompted in the nCMS to briefly inform the target person about the content of the interview, hand over the data protection information sheet, and obtain informed consent. The data protection information sheet was available in German as well as six additional languages. While it was mandatory to hand over the German version, interviewers were to provide a foreign-language version at their discretion if the target person appeared to have language problems. If a target person declined to give consent, the interviewer left and recorded a specific nonresponse code in the nCMS.

Interviewers were also required to audio record at least two interviews in the early phase of fieldwork. The interviewers were asked to use their judgment to decide whether to attempt an audio recording. If they felt that an audio recording

was an option, they were to inform the respondent about data protection by reading out a standardized text that was shown on the nCMS screen in order to obtain consent before initiating the recording.

A central function of the nCMS was to launch the PIAAC interview. Interviewers could start the international PIAAC software in the nCMS and were automatically directed to the background questionnaire and assessment. Upon successful completion of the interview, the nCMS ensured a smooth exit from the international PIAAC software and redirected interviewers back to the nCMS environment. In cases where the interview was audio-recorded, interviewers had to stop the recording. The nCMS prompted interviewers to ask the respondent whether the recording could be kept or whether they preferred to have it deleted. If there was any indication that the respondent may have felt uncomfortable with the recording, the interviewer deleted the recording through the nCMS.

After the interview was completed, the nCMS guided the interviewer through the final steps, starting with handing over the monetary incentive to the respondent. At this time, interviewers asked respondents for their consent to provide a telephone number for quality control purposes. If respondents agreed, their phone number was recorded and used during validation back checks to confirm that the interview had actually taken place (see Section 4.4). It was also necessary to document whether or not an interpreter had been used to support the respondent during the background questionnaire, and, if so, whether the interpreter was related to the respondent or was an external person.

After leaving the respondent's home, interviewers were asked to answer a few questions regarding the interview situation (see Section 4.5.3).

Special Situations in the Main Interview

The PIAAC design and standards allowed the interview to be paused in the background questionnaire, but not during the cognitive assessment. The nCMS had to accommodate this special situation. If the background questionnaire was paused for an extended period (e.g., several hours or days), the nCMS ensured that the session was properly closed in the international PIAAC software and that all data collected up to that point were securely saved. It also guaranteed that the interview could be resumed later without any issues, allowing interviewers to return directly to the point in the background questionnaire where they had previously paused.

The nCMS was also designed to handle the breakoff of interviews, which could occur at any time during the interview. If an interview could not be continued, interviewers were required to break it off and document the reason for the break-off. In this case, the nCMS closed the international PIAAC software and saved all data collected up to that point. Depending on where the breakoff occurred (background questionnaire versus cognitive assessment), the breakoff code lists differed and the nCMS routed the interviewer to the appropriate list.

Path Doorstep Interview

The doorstep interview is a short questionnaire that target persons with language barriers could complete directly at the doorstep using the interviewer's tablet (for more information on the doorstep interview, see Section 2.1.6). If the doorstep interview was an option, the interviewer followed the path *Doorstep Interview* in

the nCMS. As in the main interview, the nCMS guided the interviewers through the necessary steps to obtain consent before collecting the doorstep data. This included handing over the data protection information sheet in German and in another language, as well as an FAQ sheet (also available in several languages). Next, the interviewer selected one of 18 available doorstep interview languages and launched the doorstep interview in the international PIAAC software. Upon completion of the doorstep interview, the international software was closed and redirected to the nCMS. Any breakoffs to the doorstep interview were handled by the nCMS in a comparable way as for the main interview.

Path Documenting a Nonresponse Case

In cases where the target person refused to participate or could not be reached despite numerous contact attempts, and all interviewer efforts had been exhausted, a final nonresponse code had to be recorded in the nCMS. Whenever possible, interviewers were asked to provide additional information about the case, such as further comments on the reason for nonresponse or their assessment of whether and when a follow-up might be appropriate. PIAAC used a comprehensive list of disposition codes to capture various situations and contexts (see Section 4.5.1). In some cases, the reason for nonparticipation was not entirely clear, but rather a combination of different issues. A hierarchical approach that guided interviewers step-by-step through this part of the workflow was implemented in the nCMS.

Data Transmission

During the data collection period, interviewers were expected to transmit PIAAC data (from the nCMS, the international PIAAC software, and the audio recordings) to the field agency at regular intervals – ideally on a daily basis. The data from the nCMS were extensively used for monitoring, quality control, and reporting. In addition, the nCMS allowed the field agency to assign cases or reassign cases to interviewers when a replacement was necessary and to remotely delete all data from an interviewer's tablet.

3.5 Hardware and Technical Implementation

The complexity of the PIAAC survey was also reflected in the international information technology specifications (OECD, 2022). Specifications with respect to hardware, operating system, and configuration were defined in order to ensure that all components of the international PIAAC software would run smoothly on a tablet computer. The international PIAAC software – technically designated as the PIAAC Delivery System – was programmed by the international Consortium. Countries were responsible for integrating the international software into their national management systems, in accordance with the software interface standards. The national IT Coordinators played a crucial role in all technical preparations: They served as liaison between the Consortium, the National Centre, and the field agency, provided technical support to the field agency (e.g., for the integration process, selection of appropriate hardware, first-level support for interviewers during fieldwork), and were responsible for quality control and technical docu-

mentation. The national IT Coordination in Germany was handled by the DIPF | Leibniz Institute for Research and Information in Education.

Interviewer Tablets

In order to meet the hardware requirements specified by the international Consortium for the interviewer tablets, the field agency purchased new hardware for PIAAC Germany Cycle 2. The selection process involved several iterations, and various options were preordered for a more direct and thorough evaluation. The Lenovo X390 Yoga device was chosen, as it met or exceeded all international requirements and was approved by the international Consortium. Only one international recommendation was not followed: In Germany, a glossy screen was favoured over the recommended non-glare screen because the presentation of items was qualitatively superior compared to the non-glare screens which were evaluated in the selection process.

The Lenovo X390 Yoga is a 2-in-1 laptop, also called convertible laptop²² and includes tablet and laptop features (e.g., both a keyboard and a touchscreen). It could be transferred from laptop mode – which was used during the computer-assisted administration of the background questionnaire – to tablet mode – which was used during the self-administered cognitive assessment and the doorstep interview – by rotating the keyboard to the back of the screen; in the tablet mode, the keyboard was inactive. The most important features of the Lenovo X390 Yoga are:

- Display Type 13.3” Full HD Touch Screen
- Processor Intel® Core™ i7-8565U Processor
- Memory 16GB
- Hard Drive 512GB Solid State Drive
- Stylus ThinkPad Pen Pro (Garaged)
- Up to 14.5 hours of battery life
- Weight: about 1.29 kg

The field agency purchased 130 devices prior to the field trial. Because of the one-year delay due to the COVID-19 pandemic, the PIAAC tablets were three years old at the time of the main study. All tablets were therefore checked and, as a precaution, equipped with new original batteries for the main study.

The PIAAC hardware standards were not only important for the functionality of the international PIAAC software, but also for standardizing the display of assessment items across countries – an important component of comparability. Standards for display size and aspect ratio were therefore essential. In PIAAC Cycle 2, the aspect ratio for the display was 16:9. This differed from the standard in Cycle 1 (4:3), but a change was unavoidable because the displays used in Cycle 1 were no longer available.²³

As required for the cognitive assessment, the Lenovo X390 Yoga was equipped with an active stylus, which was stored in a dedicated compartment within the

22 Although it is technically a convertible laptop, the terminology PIAAC tablet has become established and is used in this report.

23 To remain comparable to Cycle 1, the item design remained unchanged in Cycle 2, but the item display was centered with narrow white areas on the left and right side.

keyboard section. The stylus battery recharged automatically when stored. In addition, a second, larger active stylus with a conventional battery was purchased separately (as an alternative option or to be used if the built-in stylus battery ran low).

PIAAC Software

The international Consortium developed a specific survey software application for PIAAC Cycle 2, the *PIAAC Delivery System (PDS)*. The PDS encapsulated the workflow for the administration of the PIAAC interview – the background questionnaire and the cognitive assessment as well as the doorstep interview.²⁴ Each country was provided with their specific national PDS, with the background questionnaire and cognitive assessment in the nationally translated and adapted versions. Unique national keys ensured that the PDS could only be accessed by authorized users. The PDS was run locally on the interviewer tablet and certain hardware and configuration requirements needed to be fulfilled to ensure a stable and largely error-free instrument delivery.

The PDS for Germany was integrated into the nCMS by the field agency with the support of the national IT Coordinator. For this purpose, the international Consortium equipped the PDS with an Application Programming Interface (API), which covered functionalities such as starting or resuming a case, storing data, importing and recovering data, as well as applying PDS software patches. The nCMS launched, controlled and terminated the PDS via the API. All components – the national PDS, the API, and nCMS – as well as the integrated system were rigorously tested for each release of the PDS, both prior to the field trial and the main study. The background questionnaire and the cognitive assessment were tested both stand-alone and in the integrated technical environment (see Section 2.3). Testing was carried out by the National Centre, the national IT Coordinator, and the field agency, with different foci.

Each PIAAC interviewer was provided with their own PIAAC tablet for use solely in the PIAAC data collection. Windows 10 Professional was installed on all PIAAC tablets with integrated antivirus software and HDD (Hard Disk Drive) encryption. The PDS was installed following the guidelines provided by the Consortium. All tablets were configured according to the comprehensive international and national specifications as provided by the national IT Coordinator. This included, for example, setting the device to desktop mode, adjusting brightness and microphone volume (relevant for the audio recordings) to specific settings, or deactivating unnecessary or unwanted features (e.g., Bluetooth, certain touchpad functionalities, computer mouse). The PIAAC tablets were exclusively equipped with software required for the PIAAC data collection. The remote access software TeamViewer was also installed on all laptops. It allowed the technical support team at the field agency to access interviewer tablets via a secure LTE connection to fix software problems and apply software patches. To harmonize the software on all laptops, a so-called software image containing all software and settings was prepared and exhaustively tested before it was rolled out. This ensured that the software was exactly the same across all interviewer laptops. Data transfers were realized via the encrypted LTE connection to the field agency.

²⁴ Whereas in Cycle 1 the encapsulation was realized via a virtual machine, Cycle 2 used an Electron application.

Technical documentation was handled using a ticket system (Redmine), which was used to manage document and feature requests, as well as software and hardware issues during testing. The ticket system was available during fieldwork to handle technical fieldwork support and document issues encountered. There were, however, only few technical problems during data collection.

4

Implementation of Fieldwork

The design of PIAAC in Cycle 2 foresees an eight-month data collection phase for the main study, with the target of achieving between 4 000 to 5 000 completed cases.²⁵ As described in the introduction to this report, the PIAAC Cycle 2 timeline was postponed due to the COVID-19 pandemic. Under the new timeline, the main study was scheduled for September 1, 2022 to April 30, 2023.

In Germany, data collection for the main study began on September 1, 2022 and continued until May 31, 2023. The extension to a nine-month data collection period was due to challenges encountered during fieldwork, most notably the slower than expected progress in data collection. Given Germany's two-stage sampling design, the minimum target was set at 4 500 completed cases. Nonetheless, the initial national goal aimed for 5 000 completed cases out of a gross sample of 11 200 selected individuals across 320 sample points, with 35 addresses per point.

Fieldwork for a complex survey such as PIAAC is challenging in itself. Guidelines for fieldwork have become more demanding and, over the last few years, privacy issues have led to greater restrictions. Over and beyond these difficulties, unexpected problems during fieldwork led to the revision of carefully laid plans, sometimes multiple times. This required a high degree of flexibility from all parties involved. Despite these issues, fieldwork results and results from interview validation show that data collection in Germany was overall successful.

4.1 Field Period Organization and Sample Releases

A long data collection period over several months needs careful planning and monitoring. The field agency developed the following initial fieldwork plan:

- Conduct fieldwork in several phases (including reissue phases), with a core workload planned from September to November 2022. Aim to achieve approximately 300 interviews per week between October and mid-November, with a weekly production of around 200–250 interviews before and after this period. The main phase was expected to last 22 weeks, concluding in mid- to late February 2023, with a target of approximately 4 500 interviews (or around 90 % of the total target). On average, this would have corresponded to 200 interviews per week.
- Release sample points step by step to enable interviewers to work efficiently and effectively.
- Ensure sufficient time for difficult or hard-to-reach cases, as well as for tracing and recontacting target persons who have moved to new addresses.
- Instruct interviewers to begin concentrated fieldwork efforts immediately after the in-person trainings in order to consolidate their newly acquired knowledge

²⁵ The target number of cases depends on the national sampling design. For the definition of a completed case (which is, for the most part, an interview), see Section 4.5.

and benefit from the motivational momentum. To support this, avoid or reduce additional workload from other surveys.

- Coordinate the mailing of the invitation package (advance letter, information materials, and the prepaid incentive). The mailing should arrive shortly before the interviewer makes initial contact, ensuring it is fresh in the target person's mind at the time of first contact.
- Conduct a reissue phase from mid-February to the end of April 2023, with a goal of 500 interviews. Instruct interviewers to thoroughly document all reasons for non-interviews during the main phase – particularly refusals – to enable the field agency and the National Centre to assess the eligibility of cases for follow-up activities.

In August 2022, prior to the official start of data collection, it became apparent that the original plans could not be maintained due to a shortage of available interviewer capacities. A significant number of interviewers were committed to other projects and had no resources to work on PIAAC. As a result, only a small portion of the sample could be released to the field at the beginning of September. The initial goal was to complete a large proportion of the cases in 2022, allowing sufficient time for follow-up activities. However, this objective could not be met for several reasons: Interviewers remained engaged with other studies, some areas (sample points) were insufficiently staffed, and some interviewers withdrew temporarily from their assignments due to illness or other personal reasons.

Consequently, only a step-by-step release of sample points and cases was possible, with the goal of maintaining a balanced workload for interviewers. In the end, ten waves were implemented for the main working phase (see Table 4.1), each with a different number of sample points and cases. The number of sample points released in each wave was determined by interviewer availability and the development and status of data collection. The first wave started on September 1, 2022 and the tenth wave on February 9, 2023.

Table 4.1: Waves and Number of Cases of Main Working Phase

Wave	Field start	Number of cases
1	1 Sep 2022	1 890
2	15 Sep 2022	2 345
3	29 Sep 2022	2 065
4	13 Oct 2022	840
5	27 Oct 2022	350
6	10 Nov 2022	525
7	5 Dec 2022	385
8	9 Jan 2023	1 260
9	19 Jan 2023	490
10	9 Feb 2023	1 050

Although a strict end date had initially been planned for each wave, it proved impractical to enforce. This was mainly because, at the time a new wave was released, some interviewers still had unworked or partly worked cases from previous waves or had scheduled appointments with target persons beyond the theoretical end of a wave. Closing a wave only to reopen cases in a new wave would have been inefficient and counterproductive. Therefore, a uniform end date for fieldwork was defined for the entire main working phase (May 31, 2023). In practice, interviews were still conducted for seven of the ten waves of the main working phase in May 2023.

Following the main working phase, a reissue phase was implemented, comprising seven waves (see Table 4.2). The objective of this phase was to process all non-final cases in accordance with the contacting requirements (see below for further details), and to provide all target persons with an opportunity to participate, even if they had previously not responded for a variety of reasons (e.g., extended absence). In preparation for the reissue phase, all information collected by interviewers on nonresponse cases was carefully reviewed. This included a critical assessment of the disposition codes assigned by interviewers. Different follow-up letters were sent to all cases that were identified for the reissue phase. The reissue phase is described in more detail in Section 4.3.4.

Table 4.2: Waves and Number of Cases of Reissue Phase

Wave	Field start	Number of cases
21	23 Feb 2023	396
	28 Mar 2023	399
	4 May 2023	144
22	23 Feb 2023	284
23	23 Feb 2023	767
24	28 Mar 2023	92
25	28 Mar 2023	1 209
26	4 May 2023	11
27	4 May 2023	154

A total of 939 cases were assigned to reissue wave 21. Follow-up letters were sent on three separate dates in February, March, and May 2023 (see Table 4.2 for exact dates). Three reissue waves (22, 24, and 26) were based on newly researched addresses. In these cases, interviewers had previously coded address-related issues (e.g., incorrect address, respondent no longer at address) as the final disposition. The field agency then attempted to research new addresses. As a result, 387 newly researched addresses were fielded in 2023. Finally, a large group of other nonresponse cases ($n = 2\,130$), including, for example, non-contacts, absences, or short-term illnesses, were grouped and reissued in waves 23, 25, and 27 between late February and early May 2023. As with the main working phase, all waves of the reissue phase were closed on May 31, 2023.

4.2 Contacting Rules and Disposition Codes

A well-designed contact strategy was essential to ensure that interviewers made every effort to reach target individuals. The international Consortium prescribed best-practice standards for this purpose and also defined a set of fieldwork outcomes – *disposition codes* – to be used consistently across countries (OECD, 2022).

4.2.1 Contacting Rules

According to the PIAAC standards, the first contact with the target person was to be made in person, at the respondent's home (OECD, 2022). At least six personal contact attempts were required before a case could be closed. These attempts had to be made on different days and at different times, including at least one attempt on a weekday, one on a weekend, and one in the evening during the week. The field agency was responsible for ensuring and monitoring interviewers' compliance with these requirements and had to make every effort to minimize the overall non-contact rate. The efforts were successful: In Germany, an acceptable non-contact rate of 2.5 % was achieved.

Interviewers were required to document all contact attempts and actual contacts in the national Case Management System (nCMS, see also Section 3.4) and were instructed to regularly transfer their data to home office, ideally on a daily basis. An up-to-date overview of the fieldwork status is crucial for monitoring and informs decisions to optimize data collection processes. Contact documentation included the following information for each contact or contact attempt:

- *day of the week, date and time of contact*
- *type of contact*: in person; via intercom; by phone – either a call made by the interviewer or received from the target person; via other means such as SMS or WhatsApp messages; information provided by home office indicating that the respondent had called and provided feedback, such as a refusal
- *contact outcome codes*: appointment with target person; contact with target person – no appointment; appointment with contact person; contact with contact person – no appointment; target person currently sick or unable to participate; target/contact person does not speak sufficient German; someone at home but door not opened; not at home; target person has moved; no access to building/apartment
- *additional information (if applicable)*: new respondent address; personal notes related to the case or address

The nCMS also allowed interviewers to record an appointment date for the interview. Interviewers could select the most appropriate code describing the on-site situation from the list of predefined contact codes, which all reflected an interim case status and required the interviewer to make a follow-up visit. In cases where the target person had moved and interviewers were able to determine their new address, provided it was within their radius of work, they continued their contact attempts there. If they were unable to obtain the new address, or if they could not reach and work the new address themselves, they returned the case to home office.

Home office then either initiated an address search or reassigned the new address to another interviewer.

4.2.2 Disposition Codes and Documentation of Reasons for Nonresponse

The international Consortium defined a set of disposition codes corresponding to various outcomes and stages/modules of the data collection process, including (OECD, 2022, pp. 133–135):

- DISP_CIBQ: case initialization and background questionnaire
- DISP_DS: doorstep interview
- DISP_TUT, DISP_LOC, DISP_CMP, DISP_CBA, and DISP_MAIN: separate parts of the assessments (same set of codes used for each variable)

These final outcomes could be adapted to national requirements, for example, to provide interviewers with codes they were already familiar with or to reflect specific national circumstances. However, all national adaptations had to be mapped onto the international set of disposition codes. A completed interview, consisting of both a completed background questionnaire and a completed assessment, was assigned code 1 in each of the variables DISP_CIBQ, DISP_TUT, DISP_LOC, DISP_CMP,²⁶ DISP_CBA, and DISP_MAIN.

The key disposition code variable used for all cases of the gross sample, regardless of status (i.e., main interview, doorstep interview, breakoff, or nonresponse) was the DISP_CIBQ for case initialisation. This disposition code differentiates interviews, partial interviews, and different types of nonresponse. Cases with doorstep interviews were always assigned code 7, as these target persons were unable to complete the background questionnaire due to language barriers. For each nonresponse case, interviewers were required to select a final disposition code. Table 4.3 shows the international list of disposition codes for DISP_CIBQ and how these were mapped onto the national disposition codes to record a final disposition at the case initialisation level.

²⁶ Only respondents who were administered the components.

Table 4.3: International PIAAC Disposition Codes and National Disposition Codes for Case Initialisation

International disposition codes (DISP_CIBQ)	National disposition codes
1 Complete	1 Completed interview
3 Partial complete/breakoff	<i>Various codes (e.g., refusal and codes analogous to 7-16 below)</i>
4 Refusal by sample person, e.g., refusal to participate due to time constraints or lack of interest	4 Refusal by sample person 66 Target person did not consent to the interview [set in data management]
5 Refusal by other person	5 Refusal by other person
7 Language barrier	7 Language problem (no or insufficient knowledge of German)
8 Reading and writing difficulty	8 Reading and writing difficulty
9 Learning/mental disability	9 Learning/mental disability
12 Hearing impairment	12 Hearing impairment
13 Blindness/visual impairment	13 Visual impairment (including blindness)
14 Speech impairment	14 Speech impairment
15 Physical disability	15 Physical disability
16 Other disability	16 Other disability
	37 Sample person sick (long-term)
17 Other (unspecified), e.g., sickness, falsification or unusual circumstances	38 Other reason or unusual circumstances, specify: ... 39 Falsification [set in data management]
18 Death	18 Death
20 Unknown whereabouts (for population registry samples only)	20 Sample person not found/not known at address
21 Maximum number of calls; i.e., sample person not successfully contacted	21 Repeatedly no contact
22 Invalid address (for population registry samples only)	22 Address untraceable/invalid address
23 Moved within the country but not successfully located by the interviewers	33 Moved within Germany, new address unknown
24 Temporarily absent/unavailable during field period	24 Absent, unavailable during field period
25 Not in target population, i.e., not in age range, or moved outside the country or into institution	35 Moved outside country 36 Sample person lives in institution
27 Duplication – already interviewed	<i>Not applicable in Germany</i>
90 Technical problem	90 Technical problem, specify: ...

Note. The national codes, based on the German terms used during fieldwork, were back translated into English and may not fully align with the international terminology.

In addition to the list of national codes shown here, there was one further code available to interviewers in Germany: *Code 34 – Moved within Germany, new address forwarded to the home office*. Code 90 was intended for use in cases when interviewers encountered technical problems when starting or administering the background questionnaire. However, it was rarely needed and typically assigned during data management.

The documentation of nonresponse followed a multi-stage process (implemented in the nCMS). At the first stage, interviewers were required to indicate which of the following categories best captured the reason for nonresponse (multiple selections allowed):

- 1) language barrier or disability
- 2) address-related issue
- 3) refusal
- 4) other reason

In most cases, only one category was applicable. Based on the selected category, different sets of questions followed. A routine was implemented in the nCMS to ensure that, if more than one reason was selected, the follow-up screen would always display the list of additional, more specific reasons corresponding to the nonresponse code with the highest priority (i.e., the lowest rank). For example, if a respondent had both a disability (code 1) and refused (code 3), a more detailed list of reasons for language barriers and disabilities was displayed, but not the list of reasons for refusals.

After selecting “language barrier or disability”, interviewers could choose from several disability-related options or select a language barrier code, for example in cases where communication with the target person was not possible and a door-step interview could not be offered at all.

If code 2 – an address-related issue – was the highest-priority selection at the first stage, interviewers were required to specify the type of address problem. In many cases, however, interviewers needed to select code 3 – refusal – as the highest-priority option at the first stage. In the subsequent stage, they had to indicate whether the target person had personally refused (code 4) or whether another person refused for the target person (code 5; see Table 4.3). In all cases of refusal, interviewers were required to specify the reason(s) and provide their personal evaluation as to whether the target person could be recontacted during the reissue phase. This evaluation was crucial, as certain reasons permitted a recontact attempt (*soft refusal*), whereas others strictly prohibited any further contact (*hard refusal*). Interviewers could choose multiple reasons from two separate lists, depending on whether the refusal came from the target person or some other person. These lists included reasons such as “Currently not willing to participate but may be contacted at later point in time” or “Target person raised data protection concerns”. The field agency and the National Centre continuously reviewed the selected options and free-text entries during the field period and jointly decided whether cases were eligible for follow-up or not.

Lastly, if the reasons for nonresponse were not a disability, a language barrier, an address-related issue, or a refusal, interviewers could choose option 4, another reason, at the first stage. The nonresponse codes 18, 21, 24, 37, or 38 (see Table 4.3) could subsequently be selected. *Code 21 – Repeatedly no contact* was to be

selected only if no one was ever reached at the address. If the previously described disposition codes were not appropriate, interviewers were asked to describe the specific circumstances (free-text entry). During data management, the field agency and the National Centre reviewed these free-text entries and determined whether recoding into a more appropriate code was possible. If not, the “other” code remained as the final one.

Respondents broke off the interview in very few cases. Depending on the stage at which the interview was terminated, interviewers could select the appropriate break-off code from different lists (displayed in the nCMS). These lists included codes indicating whether the breakoff was due to a refusal, a technical problem, a disability, or another reason. The code “Refusal to continue the interview” was to be used only if none of the other break-off reasons applied. That is, the respondent was physically and mentally able to participate and had the necessary skills, but simply no longer wished to continue with the interview. If the respondent had to break off the interview due to an unforeseeable occurrence, the specific circumstance was to be documented. For the assessment, an additional break-off code was available, indicating “Insufficient (or lacking) tablet skills”.

4.3 Fieldwork Monitoring and Follow-Up Strategies

A key objective of the PIAAC Cycle 2 main study was to ensure a high-quality data collection in full adherence with the PIAAC survey protocols, and in this way to also achieve maximum comparability across countries and interviewers. At the international level, measures of quality assurance and quality control were set up that checked and supported compliance with the PIAAC standards and procedures and enabled a systematic monitoring of fieldwork progress, including early flagging of potential problems (Thornton et al., 2025).

Monitoring the data collection was, however, primarily a national responsibility. While international standards and processes shaped the national monitoring work, standard national procedures were already established both at the field agency and at the National Centre. Thus, national monitoring was a fruitful mixture of international and national approaches.

4.3.1 International Fieldwork Monitoring

A detailed overview of international quality control activities can be found in Thornton et al. (2025). The international Consortium developed a National Survey Design and Planning Report in which countries were required to document their specific plans for meeting the PIAAC standards and guidelines in the main study. This structure was helpful for countries as it allowed them to lay out and specify their national plans and fine-tune their processes after receiving international feedback. For the international Consortium, this was an important tool for cross-country quality control.

In addition to the PIAAC Standards and Guidelines, the international Consortium also produced manuals and provided training on survey operations. A Data Collection Form and a Sample Monitoring Form were completed by coun-

tries on a monthly basis during data collection. This reporting was accompanied by bi-monthly calls between the international Consortium and each National Centre, during which the key collection indicators and fieldwork projections were reviewed and evaluated. If necessary, possible interventions were addressed. The international quality control also required countries to submit a summary interviewer training report and an interviewer debriefing summary report. The German National Centre fulfilled all international quality control requirements for the main study data collection.

4.3.2 National Fieldwork Monitoring

At the national level, fieldwork monitoring was carried out by both the field agency and the National Centre. Before the data collection started, a set of key performance indicators and other pertinent information were specified for national monitoring.²⁷ During data collection, interviewers transmitted all new data (incl. interviews, contacts, refusals, and paradata) to home office, ideally on a daily basis. These data were stored in a consolidated database, from which various reports with different levels of aggregation were generated.

The field agency submitted weekly monitoring reports to the National Centre. These reports included information on the overall fieldwork status (e.g., number of interviews, doorsteps, breakoffs, refusals and other non-interviews, cases in work), overall progress, and results per fieldwork wave. They also included detailed interviewer-level indicators (including number of interviews, number of interviews on same day, interval between interviews, interview start time, duration of interview/background questionnaire/assessment, number of doorstep interviews, refusals, and not worked cases). The field agency submitted regular updates on their monitoring of validation activities (e.g., validation of specific dispositions, number of interviews flagged during telephone/postal validation, audio recordings).

The National Centre regularly reviewed the field agency's monitoring results, focusing on the following indicators (overall and at the interviewer level):

- development of PIAAC response rates and completed cases, including doorstep interviews
- distribution of disposition codes
- development of sample structure
- comparison with Microcensus data
- review of survey data quality indicators
- review of interview administration
- review of audio recordings with a focus on standardization but also on providing interviewers with as much individual feedback as possible²⁸
- monitoring of doorstep interviews, especially to ensure its proper use (not as a shortcut)

²⁷ The international Consortium offered an international dashboard for monitoring purposes. However, it did not add value in Germany. All dashboard indicators were already collected as a part of the national monitoring processes.

²⁸ Interviewers had previously requested more individual feedback, and the National Centre included positive feedback, which had a motivational effect.

The National Centre conducted weekly calls with the field agency to discuss fieldwork progress, monitoring, and quality control. These calls reviewed overall progress and production, for example, current yield, interview and refusal rates, and a comparison of the fieldwork results with projections. The performance of interviewers (individual level) was discussed, including their yield (response/cooperation rate, number of interviews, number of doorstep interviews), overall activity and contacting behaviour, and ratio of soft-to-hard refusals. The discussions aimed at deriving viable measures to improve fieldwork progress, motivate interviewers, and determine appropriate adaptive measures to be implemented in the refusal conversion/reissue phase. Another important point of discussion was the status of address-related problems and subsequent address searches. Address-related dispositions (e.g., incorrect address, target person moved) were initially followed up by the interviewers on-site. If this did not yield any useful information, the field agency recontacted the registries to try to obtain up-to-date information. If new address information was available, the case was released and had the same potential as a fresh case. Finally, evaluation of the quality of interview administration and interviewer validation were ongoing activities which were also discussed at these as well as separate meetings.

The field agency organized debriefings during fieldwork to better understand the situation in the field. A relatively early debriefing during fieldwork allayed concerns about a potential accumulation of refusals due to COVID-19. At the same time, it revealed serious difficulties with the excessive length of the interviews, which impacted not only the interviews themselves, but also the recruitment of target persons for the PIAAC interviews. Interviewer feedback also indicated that it was more difficult to reach and contact target persons than it had been during pre-COVID-19 fieldwork. The analyses of the audio recordings also provided insights into how smoothly the administration of the PIAAC interview was going. For example, glitches in the transition from the background questionnaire to the assessment were identified early on and interviewers were sent clarifying instructions which resolved the issue.

4.3.3 Major Problems Affecting Fieldwork and Procedural Adjustments

A number of unexpected problems and setbacks arose during fieldwork. The plan had been for interviewers to start working directly after interviewer training and use the motivational push from interviewer training to achieve a strong fieldwork start. However, fieldwork did not get off to the anticipated strong start and remained slow for months. Because of the COVID-19 pandemic, there was a backlog of other surveys that tied up the field agency's interviewer resources and disrupted the fieldwork of subsequent studies, including PIAAC. This added to the problem of overall significantly reduced staff capacities, which were also a consequence of the pandemic, and had a negative impact on the survey landscape in Germany. In this situation, the field agency was unable to deliver on prioritizing the PIAAC survey in the initial phase of fieldwork. This was partly due to the field agency prioritizing other surveys, for example because they had been fielded earlier and needed to be wrapped up, but also due to freelance interviewers (who often worked for various field agencies) shifting their priorities according to their

own individual constraints. The unsuccessful fieldwork start also had other repercussions. Due to the lag between the interviewer training and interviewers actually starting their work, consolidation of the training content was not as effective as planned. In addition, the invitation packages were sent out as per the original fieldwork plan. Contrary to the intention, there were often significant delays between these mailings and the first contact attempts made by interviewers. Thus, the desired positive effect of the invitation mailings and specifically also of the unconditional incentive was compromised. Even when interviewers did work on PIAAC, they often did not work cases as intensively as required, due to the tense and overstretched fieldwork constellation. In addition, compared to pre-COVID times, more interviewers were unable to work due to illness.

As described in Section 2.2.4, it became clear very early on during fieldwork that the PIAAC interview was inordinately long. This demotivated not only the respondents, but also the interviewers, and increased the logistical difficulties. Although an error in the international PIAAC software was resolved by applying a software patch to reduce the length of the assessment, it took quite some time until interviewers regained their trust in the instrument, and they often continued to factor in more time for interviews which was an additional burden to their work.

As fieldwork progressed, the field agency attempted several fieldwork “resets”, introducing new projections and binding targets. Unfortunately, these were not very successful. The field agency and the National Centre (repeatedly) discussed and updated different scenarios and possible measures and set up a timeline for crucial decisions.

The following key interventions were ultimately undertaken:

- *Flat increase in interviewer remuneration*: This measure was implemented early in fieldwork. It reflected the excess length of the PIAAC interview and was also intended to increase interviewers’ motivation to work on PIAAC assignments.
- *PIAAC bonus for employed interviewers*: Both general bonuses as well as specific bonuses for high performers were introduced for employed interviewers as of December 2022.
- *Individual remuneration arrangements*: Specific payment agreements were possible on an individual basis, for example for travelling interviewers or for interviewers with distant sample points. As the data collection period progressed, these agreements were increasingly used as a tool to boost interviewers’ efforts on PIAAC and improve fieldwork results, especially during the reissue phase.
- *Adjustment of sample size target goal*: After lengthy consultation between all parties, including the Federal Ministry of Education and Research, it was agreed that the national target of 5 000 completed cases should be set aside and that the target should instead be set at (at least) 4 500 completed cases, as required by the PIAAC Standards and Guidelines.
- *Extension of data collection period*: Data collection was extended by one month, that is, it was prolonged until the end of May 2023.

Releasing a reserve sample is a straightforward way to increase the realized sample size (and reach sample size targets), but it can have an adverse impact on response rate and sample quality. Given the sub-optimal fieldwork progress, the possible

release of a reserve sample was very carefully considered and discussed between the field agency and the National Centre as well as with the international Consortium and the Federal Ministry of Education and Research. The final decision was *not* to release a reserve sample, but to adjust the sample size goal and extend the data collection period.

Additional measures to increase interviewer commitment and motivate them to work more intensely on their PIAAC caseload during the fieldwork period included online debriefing meetings with interviewers, an end-of-year information letter, and more intensive contact between the home office and interviewers. In addition, interviewers were equipped with two new seasonal cards for their contacting activities.

4.3.4 Refusal Conversion and Reissues

Refusals need to be handled with care. In Germany, hard refusals cannot be recontacted at all (by law), neither for refusal conversion, nor for interviewer validation. At the same time, refusal conversion is essential to achieve high response rates and to improve the structural quality of the sample. It was therefore necessary to find a balance between respecting the wishes and privacy of respondents, while trying to motivate participation and convert refusals.

A conservative strategy was pursued with respect to the classification of hard and soft refusals. A definitive refusal was absolutely respected, and no further contact (or validation) was attempted. All home office refusals, that is, refusals that were logged in the hotline entry form based on target person feedback proactively reported to the field agency or the National Centre (or in very few cases the Federal Ministry of Education and Research), were regarded as hard refusals.

To maximize the number of cases eligible for reissue, while identifying those that could not be recontacted, the National Centre and the field agency jointly examined a large number of nonresponse cases from the main working phase. All possible reissues, especially refusals, were carefully evaluated from a privacy perspective based on different information sources: (1) information collected from interviewers regarding the reasons for refusal, for example *not available for interview due to lack of time* (soft refusal) or *not available for interview due to privacy reasons* (hard refusal), or whether the target person had expressly objected to being recontacted (hard refusal), (2) the interviewer's evaluation of a possible recontact, and (3) the interviewer's case notes. Many cases could be immediately excluded, as it was clear that a reissue was not permitted. Cases for which a reissue could be considered required an exhaustive review of all available notes.

For cases that were identified as reapproachable, adaptive measures were introduced in the refusal conversion/reissue phases in order to reduce nonresponse and increase survey participation in hard-to-reach groups:

- *Conditional incentives*: Respondent incentives upon completion of the PIAAC interview were increased (from 50) to 80 euros.
- *Tailored advance letters*: Four different advance letters were drafted, each targeting a different recipient group:
 - a) *Non-contacts*: This letter focused on how the recipients could proactively reach the field agency to schedule an appointment for the PIAAC interview.

- b) Non-German residents: This letter emphasized the importance of participation from people with different cultural and linguistic backgrounds. The invitation mailing included a newly developed handout with very brief information on PIAAC in eight languages.
- c) Refusals 1: This letter invited target persons to proactively contact the field agency if they changed their minds and wanted to schedule an appointment. It also confirmed that no additional contact would be made by the interviewer.
- d) Refusals 2 and other cases: This letter was relatively generic.

All letters were relatively short and highlighted the new incentive of 80 euros. They also included a QR code which led to the PIAAC homepage at GESIS.

In summary, the slow fieldwork progress meant that increased monitoring activities were of the utmost importance. The National Centre and the field agency invested considerably more time and resources than anticipated into the ongoing evaluation of fieldwork and finding solutions to the difficulties encountered.

4.4 Interview Validation

Interview validation is a key measure of quality control, subject to rigorous PIAAC standards consisting of the following central pillars (OECD, 2022): (1) verification of 10 % of each interviewer's finalized work, regardless of the fieldwork outcome, that is, including all dispositions, both interviews and non-interviews, (2) random preselection of cases to be validated (based on the gross sample), and (3) validation of 100 % of an interviewer's work if the interviewer's work is flagged as suspect. Countries were required to develop and implement appropriate validation procedures and start validation early on during fieldwork. Phone or in person validation was recommended, validation per mail was accepted. Furthermore, in order to ensure high-quality interview administration, each interviewer was required to make audio recordings of at least two interviews; these were to be reviewed and assessed for interview quality.

Standard validation practice in Germany differs from the international PIAAC validation approach in various ways. Best practice in Germany, and as such the standard procedure for the field agency, is to validate 100 % of completed interviews by directly comparing key interview data with registry data and via a postal mail questionnaire. Mail is the established mode of validation because, in general, respondents' phone numbers are not available. There are no established processes for the validation of non-interviews, as these are usually not validated.²⁹

For the second cycle of PIAAC in Germany, an elaborate validation plan was developed that went beyond the international and standard national procedures, combining both into a new, more comprehensive scheme. The foundation remained the full validation of interviews,³⁰ as the established standard national protocol for quality control. In addition, and in conformance with the

29 Because most interviewers in Germany work on a freelance basis and are paid per piece, ensuring the correct administration of the interview is the focus; interviewers have no reason to falsify non-interviews.

30 This includes completed doorstep interviews.

PIAAC requirements, cases were randomly preselected for validation in the sample file. Since this was a new and (nationally) untested procedure, it was decided to over-select cases for validation to ensure that more than 10 % of each interviewer's case load would, in the end, be validated. Thus, 20 % were preselected. The validation of non-interviews was new territory for many dispositions, and appropriate protocols were developed. As previously mentioned, some validation measures could not be implemented due to data protection regulations: In Germany, it is prohibited to recontact hard refusals, even for validation purposes. Therefore, these cases had to be excluded from direct validation (validation was only possible via desk research). With respect to soft refusals, it is preferable to try to convert refusals rather than focus on validation at the reissue phase. Therefore, refusal conversion remained the focus for these cases. The field agency also extended their processes to include telephone as an additional mode of interview validation. One of the advantages of phone validation is that it allows certain issues to be immediately clarified. A prerequisite for this mode of validation, however, is that respondents previously consent to provide their phone numbers for this purpose.

Fieldwork monitoring and validation go hand-in-hand. The focus of validation is on each interviewer's individual work to ensure that they operate according to the PIAAC survey protocols, and that no falsification (partial or full) occurs. Validation was ongoing throughout fieldwork and started immediately after a case was closed. Validation work was carried out both by the field agency and the National Centre, with somewhat different foci. It consisted of several distinct measures:

- basic validation procedures (for all interviews)
- interview validation by phone and/or postal mail
- checking non-interviews
- evaluation of audio recordings

4.4.1 Validation of Interviews

Basic validation procedures were carried out for all completed interviews. Each respondent who had completed an interview was asked to respond to a validation questionnaire which was administered either by phone or sent per mail.

Basic Validation

Basic validation procedures for interviews included checking several specific interview parameters: (a) comparison of age and gender information from official records³¹ with interview data, (b) total interview duration, duration of background questionnaire, and duration of assessment, (c) time of interview, (d) number of interviews per day, and (e) time between interviews on the same day. These basic validation checks were analyzed and monitored individually for each interviewer.

The standardized basic validation was launched after each interview. This involved comparing self-reported age and gender with registry data. Nationality was also considered as a further validation aspect. All discrepancies in gender and age/year of birth were reviewed. Deviations in age of more than 2 years were further examined at regular meetings between the field agency and the National

31 Registry data.

Centre. All available information sources were considered before coming to a conclusion: registry data, case management and interview data, information from the interviewer, and data from the validation questionnaire or calls.

Timing was checked for all interviews. The durations of the entire interview, the background questionnaire, and the assessment were determined and compared to average durations. Background questionnaires that were shorter than 20 minutes or longer than 90 minutes were flagged and checked for plausibility. This included assessing the plausibility of age or year information in response to several questions in the background questionnaire and reviewing the open-ended information. The timing data was examined for specific patterns, that is, whether certain interviewers tended to conduct very short or very long interviews.

Furthermore, the start time of interviews was reviewed. Interviews starting before 8 a.m. or after 9 p.m. were flagged and interviewers were contacted for clarification. In addition to total duration and time of day, the time between interviews was also checked. Interviewers were encouraged to organize their work efficiently and were allowed to conduct multiple interviews a day. Two interviews per day were typical, although in some cases interviewers conducted three or four interviews in a single day. For all cases in which more than one interview was conducted on a particular day, the time interval between interviews was analyzed. An interval of 45 minutes or more between interviews was regarded as acceptable. If the interval was less than 45 minutes, the travel time between addresses was estimated using Google Maps. Cases in which less than 15 minutes remained after subtracting travel time were flagged and interviewers were contacted for clarification. No falsifications were uncovered. Reasons for short time intervals included interviewers initializing a case in advance (thus launching the PIAAC software) or interviews being conducted in the same building.

Validation by Phone or Postal Mail

In addition to the basic validation procedures, all completed interviews underwent either telephone or postal validation. For this purpose, the National Centre and the field agency developed a standardized validation questionnaire based on the interviewer validation form provided by the international Consortium. The latter was adapted for national use in both telephone and postal verification. The questionnaire included nine questions, four international questions and five national extensions. The international questions asked (1) approximately how long the interview had taken, (2) whether the respondent had been asked to use the tablet device, (3) whether the interviewer had provided assistance, and, if assistance had been given (4) what type of assistance they had provided. National extensions collected information on (1) month and year of birth, (2) gender, (3) nationality, and (4) time of day the interview was conducted. The last national extension replicated a question from the PIAAC background questionnaire.

For the purpose of telephone validation, respondents were asked to provide a phone number following the completion of their interview. Respondents were informed that telephone numbers would be used solely for quality assurance, assured of compliance with data protection regulations, and asked for their consent.

For validation, all participants who provided a phone number were initially contacted by telephone. After eight unsuccessful telephone contact attempts over

a period of approximately two to three weeks, or if the number was incorrect, the cases were transferred to the postal mail validation process. All participants who did not provide a telephone number for validation purposes were contacted directly by postal mail.

Telephone validation involved a telephone interviewer calling the respondent and conducting a structured interview to confirm the identity of the respondent and to ensure the PIAAC interview had been conducted according to protocols. At the end of the interview, respondents could provide additional comments in an open-ended question. To facilitate the interview process, selected interview data (e.g., year of birth, gender, interview duration, time of interview) were preloaded into the system and onto the screen of the telephone interviewers. This allowed discrepancies in responses to be clarified immediately during the telephone interview.

In the postal mail validation process, respondents received an envelope with an introductory letter and the two-page validation questionnaire, along with a pre-paid return envelope. The letters were generally mailed on a weekly basis to ensure that the follow-up was conducted as close to the interview as possible. The postal questionnaire mirrored the telephone version in structure but, being paper-based and self-administered, did not allow for immediate clarification of inconsistencies. Data entry for the paper questionnaire was completed in a timely manner.

Table 4.4: Overview of Interview Validation per Phone and Postal Mail

Validation results	Number	Percent
Validation per phone		
Routed directly to phone validation	3 702	
Successfully validated	2 504	67.6
Validation per postal mail*		
Re-routed from phone validation	1 167	
Successfully validated	588	50.4
Directly routed to postal mail	857	
Successfully validated	473	55.2
Total Interviews	4 559	
Successfully validated	3 565	78.2

Note. Source: Data from Kantar Public (2024).

* Three interviews were erroneously not processed during validation.

Of the 4 559 respondents (completed interviews),³² 3 702 (81.2 %) provided their telephone number for validation purposes. Interview validation by phone was carried out for 2 504 respondents (68 % of those who provided a telephone number). A validation questionnaire was sent by postal mail to 1 167 respondents who could not be reached by phone. Those who declined to participate in the telephone ver-

32 Since phone numbers were only obtained from cases that completed the entire interview (background questionnaire and assessment), the figure excludes doorstep cases, cases with technical problems, and breakoffs.

ification were not contacted again. In addition, 31 respondents who could not be contacted by phone did not receive a postal mail questionnaire, as the data collection period had ended. A validation questionnaire was sent by postal mail directly to 857 respondents who had not provided a phone number. Table 4.4 shows an overview of successful validation for the two validation modes. In addition to the 3 565 interviews successfully validated through telephone or postal mail, 38 interviews were validated through audio recordings of the interview.³³ Thus, the total number of validated interviews was 3 603.

4.4.2 Validation of Doorsteps, Breakoffs, and Non-Interviews

Doorsteps and breakoffs were validated using a modified version of the interview validation procedure. A new approach was developed for the validation of non-interviews.

Validation of Doorsteps

All doorstep interviews were selected for validation. Respondents who completed a doorstep interview were not asked to provide a telephone number, as validation phone calls could not be conducted in foreign languages. Thus, postal validation questionnaires were sent out to all doorstep respondents. Each respondent received a letter and a one-page questionnaire with four questions, asking about their age, gender, whether an interviewer had visited them, and whether the interviewer had asked them to answer questions on a tablet. The cover letter and questionnaire were translated into 12 languages. When possible, respondents were sent the validation questionnaire in the language they had selected for the doorstep interview (most questionnaires were sent in English, Arabic, Polish, and Russian). Of the 118 validation questionnaires sent to doorstep respondents, 38 were returned. One of the returned questionnaires could not be used for validation purposes, as it was not completed in full, resulting in 37 successfully validated doorstep cases (31 %).

Validation of Breakoffs

Cases that were preselected for validation and resulted in a breakoff during the interview were sent a modified validation questionnaire. In total, 23 validation questionnaires were sent to respondents with breakoffs; 7 were returned (30 %) and were successfully validated.

Validation of Non-Interviews

The strategy for validating non-interviews was developed in close collaboration with the international Consortium, as this was a new and untested validation measure at the national level. The Consortium recommended setting up a home office review process to examine interviewers' notes, indicating that this could be suf-

³³ The international Consortium did not regard audio recordings as an ideal way of validating interviews, because the PIAAC requirement was only to record the background questionnaire and not the cognitive assessment. In Germany, the audio recordings included the entire interview and could therefore be used for validation.

ficient to code a successful validation. The National Centre and the field agency worked together to scrutinize all information and interviewer documentation for non-interview disposition codes in order to establish whether the allocated disposition code was consistent with the written records. After careful review and deliberation, an agreement was reached on the validation status for each randomly selected case.

The validation of refusals was particularly challenging. As previously mentioned, interviewers tried to convert refusals in the refusal conversion/reissue phase, not to validate them. *Hard refusals* could not be recontacted at all (as per data privacy legislation). Thus, the standard PIAAC recommendation to send another interviewer to validate the refusal was not possible. The only way to validate hard refusals was if interviewers provided sufficient information to assess plausibility. If a *soft refusal* was reissued to another interviewer and a second refusal was registered, it was regarded as successfully validated. The interviewer documentation was examined for cases that remained with the primary interviewer in the reissue phase. *Hotline refusals* were always considered successfully validated.³⁴ Validation of *disability codes* was not pursued actively. These codes were also subject to the team review of interviewer notes. In cases of *deaths*, the registry office was consulted, and obituary searches were carried out. The disposition *unknown whereabouts* was regarded as validated if unknown whereabouts were confirmed by the registry office or the advance letter/mailing was returned to sender (as unknown). Cases that had *moved within country or outside of country* were validated through confirmation from the registry office or by reviewing the interviewer's records. To validate *language barriers*, information on nationality from the registry was consulted (if available), and contextual information from interviewers' notes was evaluated. Attempted doorstep interviews also served to validate the language barrier disposition.

4.4.3 Monitoring Validation Procedures

During fieldwork, the field agency imported all monitoring-related data into a centralized monitoring file. This complex Excel file contained various sheets, including key metrics per individual interviewer (e.g., number of completed interviews, number of doorstep interviews), case counts and results, flagged interviewers (reasons for flag, action taken), sample information, net data from phone and postal verification, field department feedback, as well as information on audio recordings, breakoffs, and doorstep interviews. The data from both telephone and postal validation were imported into the file and reviewed by the field agency on a weekly basis. The data underwent continuous quality checks both on a case-by-case basis and aggregated at the interviewer level. The weekly reviews were aimed at detecting irregularities early on, so that corrective measures could be implemented promptly. The National Centre received anonymized data from the control file every 3–4 weeks, and all flagged cases or interviewers were discussed between the National Centre and the field agency on a regular basis to decide on further steps.

³⁴ This followed the international procedure.

Any cases involving inconsistencies were investigated using a variety of sources and methods. Deviations from registry data were sometimes resolved using responses from the background questionnaire. Short interview durations were cross-checked with relevant respondent characteristics (e.g., no work history). Rapid completion of the cognitive assessment was validated through score patterns and routing (e.g., Path 1 respondents did not receive the main assessment so that short assessment durations were plausible). For unresolved or unclear cases, follow-ups were conducted with the respective interviewers or, in some cases, with respondents via phone. If clarifications were considered to be plausible, the case was marked as validated. These comprehensive control measures ensured that timely information was available for all interviews.

The discrepancies identified through the validation questionnaires were mostly related to the following:

- *Deviations from registry data:* These were examined and found to be mostly due to data entry errors by interviewers and, rarely, due to missing registry data.
- *Interview duration:* This was not considered problematic and no clustering by interviewer was found.
- *Non-matching information on time of interview:* This was considered unproblematic as differences were mostly due to different interpretations of noon/afternoon and morning/early noon.
- *No indication that assessment was administered:* Interview metadata confirmed that cognitive tasks were completed.
- *Assistance in assessment:* This was mostly resolved in follow-up steps and could be attributed, for example, to the interviewer providing help with the background questionnaire or providing the calculator in the cognitive assessment (but not helping respondents with the assessment items).
- *Differences in responses to the PIAAC background questionnaire item:* Most of these were resolved.

During validation, three interviews with the wrong respondent were detected. These interviews were subsequently removed from the data.

4.4.4 Evaluation of Audio Recordings

A central quality control measure was the systematic evaluation of at least two audio interview recordings from each interviewer, as prescribed by the international PIAAC standards (OECD, 2022). The systematic evaluation of the audio recordings allowed to assess the adherence to the PIAAC interview protocols. The audio interviews were to be collected early during data collection, ideally around the third and tenth interview. This ensured that early intervention, if necessary, was possible and field staff could provide interviewers with feedback on required adjustments. The second recording allowed to assess the administrative quality after interviewers had gathered some experience with the PIAAC interview. It also provided an evaluation of the effectiveness of follow-up trainings, where applicable. In cases in which the second recording indicated a need for further training, a third (and in exceptional cases a fourth) recording was requested.

A total of 200 audio recordings were logged and assessed. Only very few interviewers failed to provide any recording (e.g., because they worked on PIAAC for a very short time).

In a first step, the field agency reviewed all audio recordings and anonymized them to ensure that no personal information (e.g., name, employer) was present in the recordings. The field agency also conducted a first quality assessment, flagging potential concerns before passing the recordings on to the National Centre. As per the international requirements, only the background questionnaire underwent comprehensive quality control. However, the whole interview was assessed, and possible concerns in the assessment were also flagged (e.g., interviewer provided unpermitted assistance). At the National Centre, a total of six coders coded the recordings of the background questionnaires. The coding process took place between September 2022 and April 2023. Coders coded between 6 and 88 audio recordings. Prior to their coding work, all coders received comprehensive training that covered the objectives of the quality control process, the coding scheme, and the specific procedures to be followed.

In addition to collecting formal metadata (interviewer ID, date of recording, date of coding, case ID, and coder initials), the coders evaluated the quality of the interview using a standardized set of criteria: (a) the informed consent text was read aloud and consent was given, (b) questions were not skipped, (c) questions and response options were read aloud verbatim when required, (d) interviewer correctly referred to the show cards, and (e) probing was conducted when and as necessary. Upon completion of each interview, the compliance with verbatim reading, probing behaviour, interview pacing, and the overall quality of the interview were assessed.

All coded interviews were subsequently reviewed, and a summary and formal feedback was compiled for each case. This was discussed internally in accordance with the four-eyes principle. If necessary, the field agency was asked to carry out retraining with specific interviewers, which was usually done by phone. Retraining was conducted in 49 instances and for 33 interviewers. The most common reasons for retraining were repeated shortening or reformulation of questions, answering questions for the respondent (based on prior responses), not reading all possible response options, registering incomplete employment information, not responding to clarification requests from respondents correctly (leading to wrong answers), wrongly interpreting answers (leading to incorrect routing in questionnaire), and not reading all the instructions when passing the tablet on to the respondent for the direct assessment. A video-based retraining on education was prepared by the National Centre to clarify issues that were identified during the audio recordings. Follow-up training led to improvements in the quality of interview administration.

4.4.5 Validation Results

As previously indicated, the international standards and guidelines mandated a minimum of 10 % validated cases per interviewer. Table 4.5 summarizes the percentages of validated cases per interviewer for Germany.³⁵ Over 10 % of assignments were validated for all interviewers. In fact, at least 20 % of all cases were validated for every interviewer. For the majority of the interviewers ($n = 87$), 30 % or more of their cases – including interviews, breakoffs, doorstep interviews, and non-interviews – were successfully validated. Thus, the international target was surpassed. This success can be attributed to the comprehensive national validation approach, which went beyond the international validation procedures.

Table 4.5: Percentage of Validated Cases per Interviewer

Percentage of validated cases	Number of interviewers
Less than 10 %	0
10 % to less than 20 %	0
20 % to less than 30 %	13
30 % to less than 40 %	45
40 % to less than 50 %	33
50 % and more	9
Total	100

All cases were coded according to the following internationally developed scheme which considers validation status and validation result:

- *Code 0, not selected for validation:* These cases were not subject to validation (cases not randomly preselected and with a final outcome of non-interview).
- *Code 1, selected, validated acceptably:* These cases were subject to validation and successfully validated.
- *Code 2, selected, validation unsuccessful, data deleted and not refielded:* These cases included, for example, detected falsifications.
- *Code 3, selected, validation attempt unsuccessful:* These cases were selected for validation, but the final result could not be verified (e.g., validation questionnaire not returned, insufficient information to validate the non-interview disposition code for preselected cases).
- *Code 4, selected, validation not attempted:* Cases that were randomly preselected but could not be validated (e.g., due to the end of the data collection period).

Table 4.6 shows the overall validation results for the entire gross sample and for several subgroups – including all cases with a completed background questionnaire and assessment, all doorstep interviews, all cases with a breakoff in the assessment module, and all cases with the final disposition of a non-interview.

³⁵ The results are based on the final validation results after data management and cleaning.

Table 4.6: Validation Results for Different Groups

Group	Not selected for validation		Selected, validated acceptably		Selected, validation attempt unsuccessful		Selected, validation not attempted	
	n	(%)	n	(%)	n	(%)	n	(%)
Total number of cases (n = 11 200)	5 221	(46.6)	4 221	(37.7)	1 751	(15.6)	7	(0.1)
Interview* (n = 4 559)	0	(0.0)	3 603	(79.0)	953	(20.9)	3	(0.1)
Doorstep interview (n = 118)	0	(0.0)	37	(31.4)	81	(68.6)	0	(0.0)
Breakoff** (n = 116)	93	(80.2)	7	(6.0)	16	(13.8)	0	(0.0)
Non-interview (n = 6 407)	5 128	(80.0)	574	(9.0)	701	(10.9)	4	(0.1)

* Background questionnaire and assessment completed. ** Only assessment breakoffs.

No falsifications were identified. Thus, Code 2, unsuccessfully validated, was never assigned and is not included in Table 4.6. With respect to the total number of cases (gross sample), approximately 38% of cases were validated acceptably (n = 4 221). Most of these cases (85%, n = 3 603) are interviews. This result is not surprising given the national validation strategy of validating all completed interviews. However, 574 cases with a non-interview disposition were also successfully validated, which is a novelty in the German survey landscape. Overall, out of the 6 407 cases with a non-interview disposition code, 1 279 cases were preselected for validation.

The validation of refusals was the most challenging task overall, not least due to the high number of hard refusals and the fact that this disposition could not be validated directly. Validation of hard refusals relied entirely on interviewers' documentation, which was often missing or not accurate enough.

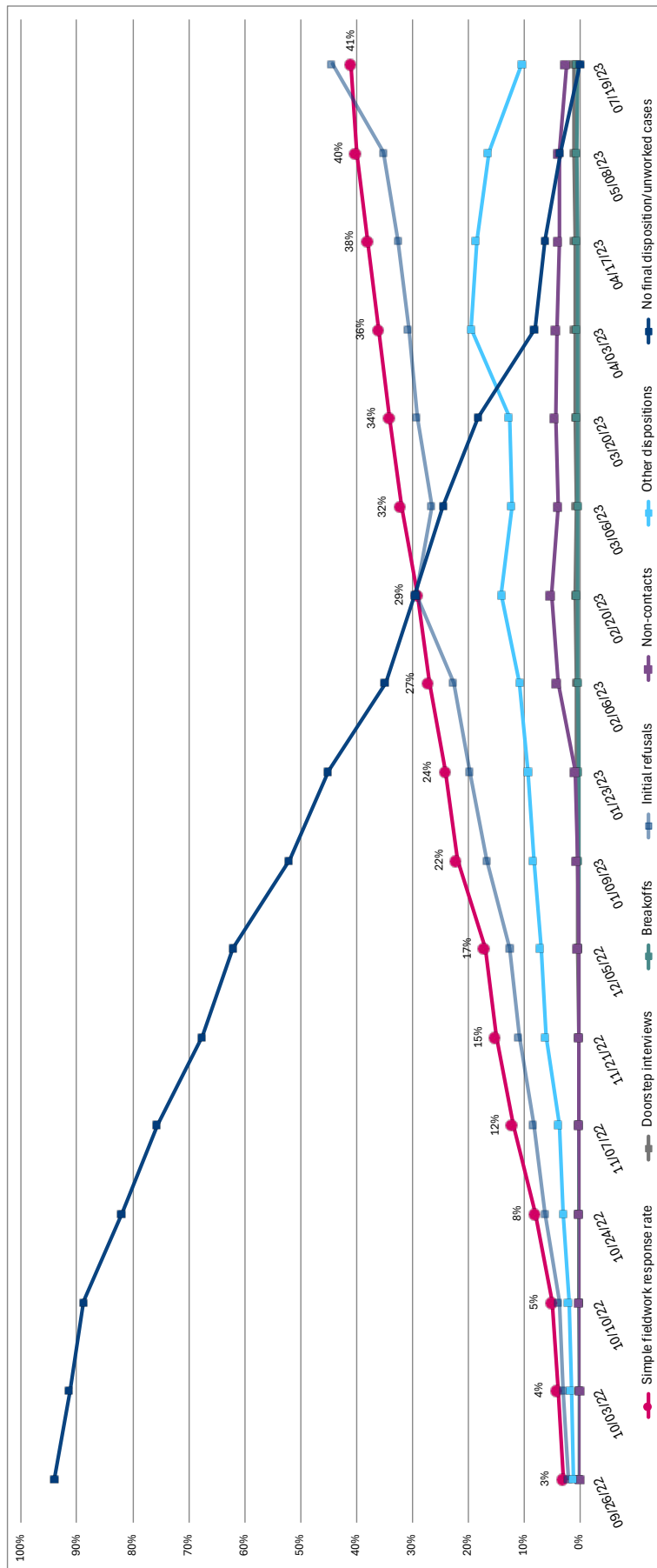
4.5 Fieldwork Results

As previously indicated, data from the nCMS and information documented by interviewers were continuously reviewed and analyzed by both the field agency and the National Centre throughout the nine-month field period in order to assess the progress of the fieldwork. Various dispositions were monitored, and a simple fieldwork response rate was computed:

$$\text{Simple fieldwork response rate} = \frac{\text{Completed interview}^{36}}{\text{Gross sample}}$$

Figure 4.1 presents an overview of the simple fieldwork response rate development, along with several (partly aggregated) outcome categories over the period from late September 2022 to mid-July 2023 (last information based on final cleaned

³⁶ Background questionnaire (DISP_CIBQ = 1,90) and assessment (DISP_MAIN = 1,7,8,9,90) completed.

Figure 4.1: Simple Fieldwork Response Rate and Fieldwork Outcomes (Cumulated Over Time)

data). The x-axis represents time (in monthly intervals), and the y-axis shows percentages. The outcome categories include: (a) doorstep interviews, (b) breakoffs, (c) initial refusals,³⁷ (d) non-contacts, (e) an aggregated category of other dispositions, such as disabilities, address-related issues, ineligible, and (f) cases without a final disposition code, for example those that had not yet been worked by interviewers at the time of observation.

The simple fieldwork response rate increased steadily over time, reaching 41 % after data cleaning. Conversely, the proportion of unworked cases declined in a nearly linear fashion throughout the observed months. The proportion of initial refusals also increased steadily, remaining slightly below the simple field response rate for most of the period, except in mid-February 2023 and in the final cleaned data, where it exceeded the response rate by a few percentage points. The proportion of non-contacts remained low for several months, indicating that interviewers persevered with their contact attempts before assigning non-contact as a final disposition. Starting in early February 2023, the non-contact rate stabilized at 4–5 %. After data cleaning, the final non-contact rate was 2.5 %. The aggregated category of other dispositions also increased steadily, peaking at nearly 20 % in early April 2023. However, this percentage was reduced to 10 % after the last two months of fieldwork, during which interviewers could revisit target persons at newly researched addresses, and after the final data cleaning process, in which the fieldwork agency and the National Centre reviewed the final dispositions from the main working and reissue phases to derive final disposition codes. The proportions of doorstep interviews and breakoffs increased gradually over time but remained at low levels.

Definition of a Completed Case in PIAAC

The international definition of a completed case in PIAAC is unique. It considers, on the one hand, the structure of the PIAAC core interview with its two parts, that is, the background questionnaire and the cognitive assessment (consisting of several modules: tutorial, locator, reading/numeracy components, and the main assessment), and, on the other hand, the doorstep interview for target persons with a language barrier. The definition also differentiates between different reasons for breakoffs in the assessment. The international Consortium defines completed cases based on the following criteria (OECD, 2022, pp. 41–42, 60):

- All cases (1) with responses to at least the key questions in the background questionnaire, including those relating to age, gender, highest level of education, employment status, and country of birth, and (2) which attempted both the tutorial and the locator module.
- All cases with technical problems during the background questionnaire or the assessment.
- Additionally, cases from the following two groups were also counted as completed cases up to a certain threshold (in Germany up to 115):³⁸

³⁷ Initial refusals in this figure represent aggregated cases over all waves and refer to initial refusals reported by the interviewer at a certain time, but it may not necessarily represent the final outcome (subsequent refusal conversion may have altered the final result).

³⁸ The number is determined by the sample design.

- Cases with responses to key background questions (see above), where the locator module in the cognitive assessment was not attempted due to one of the following reasons: lack of tablet skills, language barrier, reading or writing difficulty, learning or mental disability.
- Cases in which a doorstep interview was either successfully conducted or affected by a technical problem.

In Germany, 4 731 cases (or 42 % of the gross sample, $n = 11\,200$) were classified as *completed cases*. Of these, 4 615 cases belonged to the first group: responses were available for (at least) the key background questionnaire, and the tutorial and the locator modules were attempted. One case with a technical problem during the assessment was also included. The remaining 156 cases were capped at 115, meaning that 41 cases from the last two groups were not accepted as completed cases.

The international Consortium distinguishes between *completed cases* and *net sample cases* that received weights (OECD, 2022). To ensure that there are enough completed assessment items for the measurement of proficiencies, the Consortium set a minimum target for the number of completed cases that each country should achieve. The number of completed cases is therefore primarily of a technical nature, and countries were expected to continue collecting cases beyond this minimum threshold. The definition of a net sample case that receives weights is broader than that of a completed case and is independent of it. Ultimately, the final net sample in Germany consists of 4 793 cases, including 118 doorstep interviews, all of which were assigned a final weight.

4.5.1 PIAAC Response Rate and Final Disposition Codes

In PIAAC, the overall final response rate is derived from two components: the background questionnaire response rate and the assessment response rate. These are computed separately and then multiplied. The formulas for computing these response rates are provided in Krenzke et al. (2025, pp. 440–441) and OECD (2022, pp. 61–62).

For the computation of the *background questionnaire response rate*, the numerator included all cases with a completed background questionnaire or a completed doorstep interview, as well as cases that encountered technical problems during either the administration of the background questionnaire or the doorstep interview.

The denominator is more complex. The starting point was the total number of sampled persons (11 200). From this gross sample, several cases with specific dispositions were excluded from the response rate calculation (OECD, 2022):

- Individuals with physical disabilities, as the PIAAC interview is not designed to accommodate these circumstances.
- Individuals with language barriers but without a doorstep interview because the required doorstep language was not available, as well as individuals with reading/writing difficulties or learning/mental disabilities.
- Ineligible individuals, such as deceased persons or individuals who had moved abroad.

- Certain cases with unknown eligibility status were treated as ineligible. This included some persons who had moved away from the selected address and could not be located.
- A small number of unreachable cases (for instance due to incorrect address information), as long as the total exclusion rate (including all other exclusion reasons such as undercoverage of specific groups in the sampling frame) remained below 5 % of the target population.

The formula used to compute the background questionnaire response rate in PIAAC approximates the AAPOR response rate formula (4), but with some modifications (The American Association for Public Opinion Research, 2023, pp. 85–86). In PIAAC, the numerator includes interviews as well as *doorstep interviews* (as a PIAAC-specific element not implemented in other surveys). In contrast to AAPOR, PIAAC defines additional dispositions that are deducted from the denominator. In addition to the usual types of ineligible cases, PIAAC also considers *cases with disabilities and certain literacy-related nonrespondents as non-eligible*. Furthermore, unlike in the AAPOR computation, PIAAC also *excludes a small number of unreachable cases*. However, as in AAPOR, a proportion of cases with unknown eligibility are treated as ineligible. *Multiplying response rates from two separate modules* – the background questionnaire and the assessment module – to obtain the overall response rate is specific to PIAAC and has no counterpart in the AAPOR response rate computation.

The *assessment response rate* is defined as the number of completes divided by the number of eligibles. For the assessment, cases with completed assessments or breakoffs due to literacy-related reasons are classified as completes. Eligible respondents are defined as those who completed the background questionnaire (or experienced a technical issue during the background questionnaire), excluding individuals who broke off the assessment for specific reasons (e.g., physical disability, lack of tablet skills).

Based on the PIAAC response rate definitions, Germany achieved a background questionnaire response rate of 45.3 % and an assessment response rate of 99.1 %, which resulted in a final overall response rate of 44.9 %.³⁹

Table 4.7 presents the final distribution of the international disposition codes for Germany. Where applicable, national subcodes or other distinctive information are included. The column labeled DISP_CIBQ shows the results from the case initialisation phase (including outcome codes for non-interviews, results for partial or completed background questionnaire or doorstep interviews). The other column, labeled DISP_MAIN, displays results indicating either a completed assessment (code 1) or reasons for a breakoff in the assessment. Disposition codes that are relevant during the case initialisation but not applicable for the assessment are marked as “n/a”.

³⁹ The international Consortium reports weighted response rates for all countries in the International Technical Report (Krenzke et al., 2025). In countries with equal selection probabilities, such as Germany, the weighted and unweighted response rates are identical.

Table 4.7: Final PIAAC Disposition Codes for Germany

Final disposition codes	DISP_CIBQ		DISP_MAIN	
	n	(%)	n	(%)
1 Complete	4 675	(41.7)	4 559	(97.5)
3 Partial complete/breakoff	1	(≈ 0)	34	(0.7)
4 Refusal by sample person	4 254	(38.0)	n/a	
5 Refusal by other person	703	(6.3)	n/a	
6 Lacks the skills to use the tablet	n/a		3	(0.1)
7 Language barrier				
Total	331	(3.0)	49	(1.0)
<i>With doorstep interview</i>	118	(1.1)	n/a	
<i>Break-off doorstep interview</i>	2	(≈ 0)	n/a	
<i>Without doorstep interview</i>	211	(1.9)	n/a	
8 Reading and writing difficulty	5	(≈ 0)	6	(0.1)
9 Learning/mental disability	38	(0.3)	7	(0.1)
12 Hearing impairment	4	(≈ 0)	0	(0.0)
13 Blindness/visual impairment	4	(≈ 0)	5	(0.1)
14 Speech impairment	0	(0.0)	0	(0.0)
15 Physical disability	13	(0.1)	2	(≈ 0)
16 Other disability	27	(0.2)	1	(≈ 0)
17 Other (unspecified)				
Total	146	(1.3)	7	(0.1)
<i>37 Sample person sick (long-term)</i>	85	(0.8)	n/a	
<i>38 Other reason/unusual circumstances</i>	61	(0.5)	n/a	
18 Death	36	(0.3)	n/a	
20 Unknown whereabouts (for population registry samples only)	222	(2.0)	n/a	
21 Maximum number of calls; i.e., sample person not successfully contacted	277	(2.5)	n/a	
22 Invalid address (for population registry samples only)	20	(0.2)	n/a	
23 Moved within the country but not successfully located by the interviewers	160	(1.4)	n/a	
24 Temporarily absent/unavailable during field period	83	(0.7)	n/a	
25 Not in target population				
Total	201	(1.8)	n/a	
<i>35 Moved outside the country</i>	159	(1.4)	n/a	
<i>36 Sample person lives in institution</i>	42	(0.4)	n/a	
90 Technical problem	0	(0.0)	2	(≈ 0)
Overall total	11 200	(100.0)	4 675	(100.0)

Note. n = number of cases; n/a = not applicable.

A look at the fieldwork results shows that the primary reason for a non-interview was a refusal: 44 % of the gross sample refused to participate (codes 4 and 5), and the majority of these cases were classified as hard refusals. Other dispositions each accounted for 3 % or less of the gross sample.

Two target persons started the background questionnaire but broke it off. In one of these cases, the breakoff occurred during a later stage of the background questionnaire. As all required key information was collected, the international Consortium classified this case as a complete for the background questionnaire (DISP_CIBQ = 1). The vast majority of respondents who completed the background questionnaire also completed the assessment – almost 98 %.

In total, 115 breakoffs occurred after the background questionnaire was completed. In 20 cases, the interview was broken off after the end of the background questionnaire but before the assessment began. This was mostly due to respondents refusing to continue because of a lack of time, interest, or motivation. In addition, the tutorial module was broken off 35 times, in almost 70 % of the cases due to a language barrier. Nineteen respondents did not complete the locator module, with just over half of these breakoffs again due to a language barrier. Eight breakoffs occurred in the components module due to lack of interest, language barriers, and learning/mental disabilities. Thirty-three respondents broke off in the final module – the main assessment. Many of these respondents (51 %) refused to continue due to lack of time, interest, or motivation. In sum, motivational factors and language barriers were the most frequent reasons for breakoffs during the assessment. The lack of technical skills required to use the tablet was not an issue and was only reported on three occasions.

Overall, the 4 559 interviews with completed assessments were conducted by 99 interviewers.⁴⁰ On average, this corresponds to 46 completed interviews per interviewer. However, the number of interviews was not evenly distributed across interviewers, with a range from 2 to 134 completed interviews per interviewer.

4.5.2 Contacting Results and Behaviour

In Germany, interviewers documented a total of 48 683 temporary and final contacts and contact attempts (hereafter referred to as contacts) across all phases of the fieldwork (main working and reissue phases). On average, 10.4 contacts were needed to obtain 4 675 interviews.⁴¹ In the most extreme case, the same target person was contacted 52 times before the final outcome was reached: a non-interview. In contrast, in 13 % of all contacts, the final outcome was achieved during the first contact: 46 % resulted in a refusal, 15 % had other unspecified reasons, and 14 % completed the interview. For about one-third (33 %) of the gross sample, five or more contacts were needed to obtain a final outcome.

40 Of the 100 interviewers who completed PIAAC assignments, one interviewer worked only as a contactor and did not administer any interviews.

41 In this context, an interview refers to a completed background questionnaire; whether the assessment was completed or broken off is irrelevant for evaluating contacting results.

Focusing on interviews, a total of 19 120 contacts were documented, averaging 4.1 contacts per case and with a maximum of 30 contacts. Five or more contacts were needed to obtain the interview in 29 % of these 19 120 cases.

Of the 48 683 contacts, the vast majority (75 %) were made in person. Among the remaining contact types:

- 10 % were made by the interviewer via telephone.
- 5 % involved the target person calling the interviewer.
- Around 3 % each were made via intercom or through other media (e.g., email or SMS), initiated either by the interviewer or the target person.
- In roughly 1 % of cases, target persons contacted the field agency directly, indicating that they did not wish to participate. In these instances, no contact outcome was recorded by the interviewers; instead, the contact was logged by the field agency.

As for the day of the week when contacts occurred, contact attempts were relatively evenly distributed from Monday through Friday, with 15–18 % per weekday. Thirteen percent of contacts were made on a Saturday, and only 3 % on a Sunday. Most contacts took place between 4–8 p.m. (41 %) and 12–4 p.m. (39 %), 17 % in the morning (before noon), and about 3 % after 8 p.m.

According to PIAAC standards, the first contact with the target person must be made in person (OECD, 2022). In Germany, this standard was met for the most part of the gross sample: In 96 % of the 11 200 cases, interviewers made the first contact in person, including 3 % via intercom. About 3 % of the recorded contacts were initiated by the target persons themselves (e.g., calls and messages to the field agency's hotline or directly to the interviewer). In 1 % of the cases, however, interviewers documented that they either called the target person by phone or sent them a message. This outcome does not align with the standards but is most likely the result of inaccurate documentation by interviewers. Typically, interviewers do not have access to personal contact details (such as phone numbers or email addresses) prior to making the first in-person contact with the target person.

In cases where a non-contact (DISP_CIBQ = 21) was recorded as the final disposition, the PIAAC standards required that at least six in-person contact attempts be made (OECD, 2022). These attempts had to be spread across different days of the week and time slots. Specifically, at least one attempt had to be made on a weekday (Monday to Friday) before 6 p.m., one on a weekday after 6 p.m., and one at the weekend. In Germany, 277 cases were assigned the final disposition code 21. The PIAAC standard was met in two thirds of these cases (66 %), and interviewers made six or more contact attempts before assigning this final code. In one case as many as 31 attempts were recorded. Despite close monitoring and regular reminders by the field agency regarding the PIAAC contacting standards, interviewers were unable to comply with the required number of contact attempts in one third of the cases (34 %). In around 7 % of all cases, only two or three contact attempts were made.

On average, 7.7 contact attempts were recorded per non-contacted target person; 94 % of the attempts were made in person, including 4 % via intercom. With respect to the distribution across different days and times, only about a quarter of the cases (26.7 %) met all three criteria (i.e., at least one contact attempt in each of the three time periods – weekday, weeknight, and weekend). In 44.8 % of the

cases, attempts were made in two of the required periods, while in 28.5 % of the cases, attempts were only made in one of the three time slots.

4.5.3 Post-Interview Questions

After the interview, interviewers responded to questions about the interview and the interview process. The post-interview questions encompassed fifteen closed and three open-response questions (see Appendix C). Eleven international questions focused on the interview setting (location, presence of other people, disturbances), possible assistance, and comprehension of the interview questions. Seven national extensions focused on difficulties with the technology, use of stylus and calculator (in the cognitive assessment), distractions caused by electronic devices, and health and safety protocols due to COVID-19.

Interviewers were instructed to respond to the post-interview questions as soon as possible after leaving the respondent's home while they could still correctly recall the interview and respond accurately.

Post-interview questions were regularly analyzed as a part of monitoring activities, enabling rapid interventions if difficulties in the interview process were identified. The analyses were carried out weekly in the first four weeks of data collection, and subsequently bi-weekly. This included checking response distributions as well as a qualitative analysis of responses to the three open-ended questions.

The analyses showed relatively stable response distributions over the fieldwork period. Table 4.8 presents a summary of the final results after fieldwork was completed. Interviewers were instructed to avoid the presence of third persons during the interview. However, this could not always be realised, for example, if the respondent's living situation did not allow for privacy. In approximately one fifth of the interviews, another person besides the respondent and interviewer was present during the interview. In the vast majority of cases, interviewers reported that the respondents often or very often understood the questions in the background questionnaire. Twelve percent of respondents asked for assistance with tablet functionalities and 8 % asked for assistance with the assessment. Although assistance during the assessment was not allowed, in 1 % of cases the interviewers reported assisting the respondents. Follow-ups with the interviewers revealed most of these instances pertained to clarification of technical issues during the assessment. The assessment was usually completed in respondents' homes. Interviewers reported that in 14 % of cases, the cognitive assessment was interrupted at least once. In total, 13 % of respondents complained about the length of the interview. However, the percentage decreased from 19 % to 11 % after an international software patch was implemented and the assessment shortened (see Section 2.2.4).

Table 4.8: Responses to Post-Interview Questions

Post-interview questions: interviewer responses	Percent
No-one else was present during interview	81.6
Someone else was present during background questionnaire	15.9
Someone else was present during assessment	11.7
Respondent understood questions in background questionnaire often or very often	96.4
Interviewer assisted during assessment	0.9
Respondent asked for assistance with features of tablet	12.1
Respondent asked for assistance with an item in assessment	7.7
Background noise during assessment	18.4
Assessment interrupted once or more	14.3
Respondent complained about length of assessment, overall	13.2
<i>Before patch*</i>	18.6
<i>After patch*</i>	11.3
Assessment completed in respondent's home	91.6
Respondent used handheld calculator in assessment	69.4
Respondent completed assessment mostly using stylus	77.0
Respondent had problems with functionality of international PIAAC software	4.3
Respondent completed assessment conscientiously all or almost all of the time	78.8
No distraction through electronic devices	92.8

* Software patch applied during main survey which reduced the length of the cognitive assessment (see Section 2.2.4).

Interviewers were asked to share additional remarks they had about the interview, the international PIAAC software, and the administration of COVID-19 safety protocols. The interviewers most often noted that respondents complained about the length of the interview, indicating that they became tired, and that their concentration decreased towards the end of the assessment.

4.5.4 Interviewer Debriefing

The impressions and experiences of interviewers during fieldwork can provide valuable insights into the implementation of the data collection. Feedback contextualizes the challenges and conditions of data collection, helping to identify lessons learnt for the future and possible implications for data quality. Following the PIAAC standards, feedback was sought via an interviewer debriefing questionnaire which was based on a form provided by the international Consortium, with national adaptations and extensions. The debriefing questionnaire was sent to 95

PIAAC interviewers;⁴² 89 interviewers completed the questionnaire. In addition, three in-person debriefing workshops were held with a total of 31 interviewers.

In hindsight, interviewers felt that the in-person interviewer training had prepared them very well for the PIAAC fieldwork. They found it easy to administer the background questionnaire. Some respondents had difficulties answering some of the questions (often respondents with language barriers). The questions on education were sometimes difficult (e.g., due to problems finding the correct category), especially for respondents with a migration background. In general, questions about parents were difficult to answer, in particular parental education. Some questions were too long and too repetitive.

The administration of the cognitive assessment was straightforward; most interviewers coped well with the more observational role it required. The assessment was found to be generally too long, respondents tired and had problems concentrating. Many texts were too long (literacy), and the reading load was strenuous and overwhelming. Respondents who worked on adaptive problem solving sometimes found the nature of the problem unclear and didn't always realize when the task started (after a trial run) and when it was completed. Some respondents had difficulties understanding the instructions in the assessment. There were sometimes problems reading the screen text, usually because the font was too small. The overwhelming feedback was that the assessment should be shorter and that more effort should be made to keep respondents motivated. To quote one interviewer response: *Demands made on respondents were not reasonable – it was an endurance test.* At the same time, interviewers reported that respondents were often immersed in their task and seemed to work conscientiously. Some tasks were fun.

Only very few respondents were hesitant about using the tablet, but some had issues with its functionalities or the software. Technical problems were rare.

Only 38 interviewers administered the doorstep interview; some found it very easy to administer, others less so. Some of the following problems were mentioned: specific languages not available; German version of doorstep interview missing in material; communication difficulties, hard to explain what a short doorstep interview is; distrust; refusal by interfering household members. It was suggested that the paper material be compressed and that a German version should be added (for the interviewers' own understanding) as well as versions in additional languages.

Interviewers felt well-equipped for their work. The goals for the response rates, however, were regarded as too high and unrealistic. Contacting target persons has become increasingly challenging. In the PIAAC fieldwork, target persons were often difficult to reach and access. This was partly due to the decreasing quality of the addresses, but also because target persons didn't react, cancelled appointments, and "ghosted" the interviewers. Common reasons for refusals were no time, no interest, or interview too long. COVID-19 was not as problematic as expected, and refusals due to COVID-19 were rare. Respondents sometimes asked interviewers to test themselves or to wear a mask, but such requests decreased with time. Reasons that motivated target persons to participate were interest in the study (important international study), curiosity, and the incentive.

42 The remaining interviewers could no longer be reached due to death, illness, or because they were no longer employed by the field agency.

The in-person debriefing workshops revealed a high level of commitment and enthusiasm for the PIAAC survey – overall, interviewers found it to be a very interesting study that they very much enjoyed working for.

4.6 Curbside Observations

While there is comprehensive data on PIAAC respondents, not much is known about nonrespondents. In the second cycle of PIAAC, a curbside observation instrument was introduced as a first step towards closing this gap. This module collects interviewer observations on the neighbourhood and dwelling unit in an equivalent manner for all target persons, that is to say both for respondents and nonrespondents. Interviewer curbside observations may be used to inform measures of adaptive survey design, be considered for nonresponse bias analyses, or, if the data has sufficient quality, even for nonresponse weighting adjustments (Krenzke et al., 2025).

The curbside observation module was an international option, but its implementation was recommended. For many countries this was an innovative approach. In Germany, however, curbside observations are not a new concept. Indeed, PIAAC Germany 2012, as well as various other surveys such as the European Social Survey (ESS) in Germany or the German General Social Survey (ALLBUS) have administered similar modules in the past (GESIS Datenarchiv für Sozialwissenschaften, 2018; infas Institut für angewandte Sozialwissenschaft, 2016; Zabal et al., 2014).

The PIAAC Consortium developed a curbside observation instrument that needed to be translated and adapted by countries. The international module included questions on type of dwelling unit, features restricting access to the residence, such as a locked building or gated community, and the physical condition of the residence (Krenzke et al., 2025). For the German version of this instrument, similar curbside information questions from other surveys in Germany were reviewed as a part of the adaptation process, with a focus on the questions from PIAAC Germany 2012. In addition to the international set, a question on the target person's presumed education was added as a national extension. This information was also collected in PIAAC Germany 2012 and was used to implement a responsive measure (Martin & Zabal, 2020).

The following curbside information was collected in PIAAC Germany 2023:

1. physical condition of the neighbourhood
2. type of house/dwelling unit
3. physical condition of the house/dwelling unit
4. presence of an intercom
5. problems accessing the house/dwelling unit
6. subjective assessment of the target person's education

In order to obtain equivalent information for respondents and nonrespondents, it is essential that the interviewers record their impressions *before* any personal contact is made with the target person, to ensure that responses remain uninfluenced. Thus, interviewers were instructed to form an impression of the residential

area and the dwelling unit of the target person on their *first visit* to the specified address and to complete the curbside questions before making any initial contact with the household and/or the target person.

In Germany, interviewers were instructed to enter the curbside information directly into the nCMS. When this was not possible, interviewers were asked to note the answers on the back of the address log and transfer them to the nCMS as quickly as possible after the first contact (attempt). The interviewer instructions emphasized that initial responses should not be changed in any way, in order to ensure equivalent data quality for both respondents and nonrespondents (including non-contacts). This requirement was especially relevant to the subjective assessment of education, to make sure that information obtained during the PIAAC interview did not bias the curbside measurement.

Target persons who contacted the study hotline to communicate their refusal were regarded as hard refusals and could not be contacted. Interviewers were informed that this could happen, and, in such cases, they were instructed to immediately close the case (unless it had already been closed by the home office). Consequently, no curbside information is available for these cases.

In order to ensure that the curbside module was implemented in the same way by all interviewers, its administration was explained in detail in the interviewer manual and in a separate interactive interviewer training module, which also strove to establish a common approach through discussions of examples of different areas and dwelling units.

Curbside observations are by nature subjective and may be susceptible to errors (e.g., West, 2013). They thus require careful evaluation and should be used with caution. Subjective assessments of target persons' education are especially challenging, as interviewers have little information on which to base their judgments. Consequently, their responses are essentially "best guesses".

A reliability study was conducted to evaluate the quality of the interviewer observations in Germany. For a subsample of the PIAAC sample ($n = 770$), another (non-PIAAC) interviewer was appointed to carry out a second set of observations. This second set was used to test how consistent, and thus how accurate, interviewers' observations were with each other.

For the reliability study, 23 sample points from the main PIAAC study were selected. Prior to fieldwork, interviewers for the reliability study were trained on the curbside module using the same materials as PIAAC interviewers. However, as the training was conducted online and not integrated into the multi-day training of the PIAAC main study, the preparation of the interviewers for the reliability study differed somewhat from that of the PIAAC interviewers.

The interrater reliability was computed for each of the six curbside observation questions. In Germany, it ranged between 55 % (target person's education) and 93 % (binary indicator on problems accessing the dwelling unit). Thus, the subjective assessments show considerably less agreement than the more objective assessments. Differences in curbside observations between interviewers may arise from varying levels of familiarity or prior knowledge of the sampling point. In addition, the timing of the observation – such as daylight, weather conditions, or nearby garbage collection – can influence how a location is assessed.

5

Weighting, Nonresponse Bias Analyses, and Variance Estimation

In large-scale international surveys such as PIAAC, high-quality sample data are essential for producing comparable estimates across countries and drawing valid conclusions. The international Consortium has established approaches to address challenges arising from varying response rates, to assess the potential for bias in estimates, and to ensure accurate calculation of standard errors that account for sampling variability in complex sampling designs (OECD, 2022). This chapter covers four related key areas: weighting, nonresponse bias analyses (NRBA), variance estimation, and design effects.

Weighting plays a critical role in population surveys, as not all sampled persons participate. Each respondent represents a portion of the target population. To enable valid population-level inferences across countries, weighting procedures must correct for undercoverage and nonresponse. Calibrated weights, aligned with known population benchmarks, reduce potential bias from nonresponse. Section 5.1 outlines the stepwise development of sampling weights – from design (or base) weights to final calibrated weights – and provides information on auxiliary variables used.

All countries with a response rate of below 70 % were required to conduct extended nonresponse bias analyses (OECD, 2022). Bias can be introduced if parts of the target population are excluded from the sampling frame (e.g., due to undercoverage) or if nonrespondents systematically differ from respondents. Section 5.2 presents the analyses conducted to evaluate potential bias after weighting adjustments were applied for Germany.

Section 5.3 on variance estimation illustrates how PIAAC addresses the accurate computation of standard errors of population estimates with replicate weights. Finally, Section 5.4 addresses the design effect for the three skill domain measures. The design effect quantifies the impact of the complex sampling design – particularly clustering – and weighting on data quality, relative to a simple random sample.

5.1 Weighting

The primary objective of calculating sampling weights in large-scale surveys such as PIAAC is to ensure that estimates derived from the sample can be generalized to the target population (for PIAAC individuals aged 16 to 65) in the most accurate and unbiased way possible. When weights are applied appropriately, survey results should reflect the characteristics of the population from which the sample was drawn, while minimizing – or ideally eliminating – potential bias. Sampling weights indicate how many individuals in the target population are represented by each respondent. They are designed to correct certain sources of error, including undercoverage, differential response rates across population subgroups, and systematic differences between respondents and nonrespondents.

The Technical Standards and Guidelines define standards and requirements for weighting in PIAAC (OECD, 2022). One of these standards specifies that, by default, weighting is performed by the international Consortium, as was the case for Germany.⁴³ This approach is intended to ensure consistency across countries and support international comparability. Countries were responsible for (a) identifying and providing the auxiliary variables for nonresponse weighting, which should be predictive of both response propensity and the survey outcome, (b) providing variables for the calibration step, and (c) checking the weights computed by the Consortium. Details regarding the weighting approach applied in PIAAC, both in general and specifically for Germany, are documented in the Consortium's International Technical Report (Van de Kerckhove, Alvarez-Rojas, et al., 2025).⁴⁴ The weighting approach in PIAAC was designed to achieve the following objectives:

- correct for unequal selection probabilities (note: not applicable in Germany, as all target persons had equal selection probabilities by design)
- compensate for undercoverage due to exclusion of certain parts of the target population from the sampling frame
- reduce the potential for nonresponse bias
- minimize sampling error by aligning the weighted sample with external population benchmarks measured with high precision

Weighting in PIAAC is conducted across several levels, with each level adjusting weights for a specific subgroup of the sample. The weighting factor from one level serves as the basis for the next level. Van de Kerckhove, Alvarez-Rojas, et al. (2025, p. 216) provide an overview of the individual weighting factors and the formulas for their computation for registry countries (such as Germany). The weighting procedure begins with the base weight, which reflects the inverse of the unit's selection probability. In Germany, all selected target persons in the gross sample have equal selection probabilities, and thus equal base weights. The subsequent weighting levels – described in more detail below – are:

- unknown eligibility adjustment
- nonresponse adjustments, split into two separate steps: non-literacy-related and literacy-related
- calibration

5.1.1 Unknown Eligibility Adjustment

In all surveys, there are cases where the eligibility status of sampled persons cannot be determined. This typically occurs when interviewers are unable to establish contact with target persons or to gather any information about the person's whereabouts. Among these cases with unknown eligibility status, some individuals are in fact ineligible (e.g., because they have moved abroad), while others are eligi-

43 Some countries conducted the weighting themselves but were required to adhere to the standards and requirements.

44 As central references for weighting, this source and the Technical Standards and Guidelines (OECD, 2022) will, in the following, only be cited in specific cases where a direct reference is necessary or helpful.

ble. However, the exact proportions remain uncertain. At the unknown eligibility adjustment level, the weights of ineligible cases are increased based on the proportion of unknowns in the gross sample. Conversely, cases with unknown eligibility are down-weighted using a factor calculated as the ratio of eligibles to all individuals with known eligibility status. In the subsequent non-literacy-related nonresponse adjustment step, these cases are classified as eligible nonrespondents. To calculate the unknown eligibility adjustment factors for subgroups with different response rates, the Consortium carried out a classification tree analysis. A classification tree produces nodes (or, in the context of weighting, cells), which group together sample persons who are homogeneous with respect to survey participation. The German National Centre supported this analysis by supplying a set of variables, specifically gender, age, citizenship, region, and municipality size, which were identified as auxiliary variables for the non-literacy-related adjustment step (further details see below). The unknown eligibility adjustment was implemented within the same weighting cells that were formed for the nonresponse adjustment.

5.1.2 Nonresponse Adjustment

In PIAAC, various reasons for nonresponse are recorded and grouped into two main categories for the purpose of nonresponse adjustment. Since language – specifically the test language – is essential for completing the tasks in the skills assessment module, literacy-related reasons for non-participation are of particular relevance in PIAAC. Thus, the nonresponse adjustment step consists of separate weighting adjustments: one for the majority of nonrespondents without a literacy-related reason, and one for those with a literacy-related reason (Van de Kerckhove, Alvarez-Rojas, et al., 2025).

For the non-literacy-related nonresponse adjustment step, it is necessary to identify auxiliary variables that are associated both with the PIAAC outcome (i.e., the assessed skills) and with survey participation. This ensures that potential bias in the data due to systematic nonresponse of certain groups can be mitigated. The variables were identified and selected by the National Centre as part of the basic NRBA. Based on these variables, the international Consortium conducted classification tree analyses to identify subgroups with varying response rates in the gross sample and to derive adjustment factors for non-literacy-related nonresponse (Krenzke et al., 2025).

In the first step, the National Centre selected candidate predictor variables that were associated with the proficiency outcomes, available for all eligible individuals in the gross sample, and met specific quality criteria (e.g., low rates of missing data, high accuracy). These variables could come from various sources, such as the sampling frame, external datasets, or data collected during fieldwork (e.g., the curbside observation module). As a minimum requirement, countries were expected to include age, gender, educational attainment, employment status, and region in the basic NRBA (Krenzke et al., 2025).

In Germany, candidate variables from three sources were investigated:

- *sampling frame*: regional information such as federal state and municipality size, as well as target person characteristics such as age, gender, and citizenship
- *curbside information*: interviewers' initial observations of a target person's living environment (e.g., type and condition of dwelling unit) and presumed educational attainment, recorded before the first contact (attempt) with target person
- *commercial database*: micro-geographic area data (e.g., at the house or street level) including economic, socio-demographic, and psychographic characteristics, such as estimates of purchasing power or the proportion of high school graduates⁴⁵

Since decisions on nonresponse adjustment variables and the basic NRBA had to be made before proficiency scores were available, the National Centre used proxy variables for proficiency in the first step. These proxy variables included highest level of schooling and professional education according to German degrees, educational attainment based on the ISCED⁴⁶ classification, and respondents' results in the locator module.

During Step 1, the National Centre applied both bivariate analyses (correlations and chi-square tests) and multivariate methods (logistic regressions) to evaluate the relationship between the proficiency proxy variables and candidate predictor variables from the various sources. Most candidate predictor variables showed significant correlations with the proficiency proxies. However, for the basic NRBA, the National Centre placed greater emphasis on the sampling frame variables, as these were of the highest quality. Age, gender, citizenship, federal state, and municipality size were selected.⁴⁷ Data from the other sources did not meet the quality standards: Data from the commercial database were too highly aggregated and had too many missing values, while the observational data collected by interviewers were too error prone.

The second step involved evaluating the relationship between the candidate predictor variables selected in Step 1 and response status. Using classification tree analysis, the international Consortium identified subgroups within the eligible gross sample that had low response rates and assessed which of the candidate predictor variables were suitable for nonresponse weighting. All five candidate predictor variables – age (continuous variable), gender (2 categories), citizenship (2 categories), federal state (14 categories), and municipality size (10 categories) – appeared in at least one node (or cell) of the resulting classification tree.⁴⁸ The classification tree output identified specific cells – defined by combinations of these variables – that showed distinct response rates. These cells served as the basis for constructing effective non-literacy-related nonresponse adjustment factors.

For the literacy-related nonresponse adjustment step, a separate classification tree analysis was carried out using the same set of variables as in the non-literacy-

45 The database was set up and distributed by microm Micromarketing-Systeme und Consult GmbH (<https://www.microm.de/>).

46 International Standard Classification of Education

47 Contrary to the requirements, employment status could not be used in this step because this variable was not available at the individual level in the sampling frame. However, employment status was taken into account during the calibration step.

48 In Cycle 1, age, citizenship, and municipality size were included in the non-literacy-related nonresponse adjustment (Zabal et al., 2014).

related nonresponse adjustment step. The key difference lies in the cases included in the analysis: Nonrespondents consisted of target persons who did not participate in PIAAC for literacy-related reasons; the respondent group comprised individuals who completed a doorstep interview as well as those who broke off the assessment due to literacy-related reasons (Van de Kerckhove, Alvarez-Rojas, et al., 2025). In Germany, the classification tree cells that were used to derive the literacy-related nonresponse adjustment factors were defined solely by federal state and citizenship.

5.1.3 Calibration

The final main weighting step in PIAAC is calibration,⁴⁹ which aligns the survey data with high-quality benchmark controls from a reliable external source. A final weight must be assigned to all units of the net sample – in Germany 4 793 cases. Age and gender are mandatory variables for calibration. All candidate variables considered for calibration need to meet the following requirements (Van de Kerckhove, Alvarez-Rojas, et al., 2025):

- *In the PIAAC data:* Variables must be available for all cases of the net sample, including those with only a doorstep interview for which only a very limited number of variables exist.
- *External data:*
 - Reliable control totals must be available for the PIAAC target population.
 - The concepts measured by the variables must correspond to those in PIAAC.
 - The external data should correspond as closely as possible to the time period of the PIAAC data collection.
 - The quality of the external data must exceed that of the PIAAC data.
 - The proportion of missing data for the variables should not exceed 5 %.⁵⁰

In Germany, the 2021 Microcensus – the most recent dataset which was accessible at the time – met the required standards and was used as the benchmark source for calibration.⁵¹ In Cycle 1, poststratification was applied, using gender, age, region, and highest general school-leaving qualification as benchmark variables (Zabal et al., 2014). For Cycle 2, methodological considerations led the National Centre to opt for raking instead of poststratification; this was approved by the international Consortium.

Raking is an iterative procedure that adjusts the marginal distributions of selected variables to the margins of the benchmark data (in this case, the Microcensus data). One key advantage of raking over poststratification is its ability to incorporate a larger number of auxiliary variables in the weighting process. For PIAAC Germany Cycle 2, raking was performed using the following dimensions

49 The following techniques were permitted (Van de Kerckhove, Alvarez-Rojas, et al., 2025): poststratification, raking, or GREG (generalized regression estimators).

50 If this condition was met but some data were still missing, countries were required to replace them using imputation (OECD, 2023a).

51 Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States of Germany, Microcensus 2021, <https://doi.org/10.21242/12211.2021.00.00.3.1.0>

(number of categories in brackets): gender by age (10), region (4), highest general school-leaving qualification (4), employment status (3), country of birth (2), and gender by employment status (4). Compared to Cycle 1, two additional variables – employment status and country of birth – were included. Both were significantly related to response behaviour and are key analytical variables.

After the initial calibration step, the international Consortium examined the data for the presence of extreme weights and applied weight trimming where necessary (for details on the trimming criterion, see Van de Kerckhove, Alvarez-Rojas, et al., 2025, p. 221). If trimming was conducted, a second calibration step followed. For Germany, it was determined that none of the weights exceeded the cutoff value, meaning that no extreme weights were present, and trimming was not required for the German data.

The international Consortium implemented a set of quality control procedures to ensure that the weighting process adhered to PIAAC standards and guidelines, and that the resulting weights were accurate and valid (Van de Kerckhove, Alvarez-Rojas, et al., 2025). In countries where the Consortium conducted the weighting, the National Centres were responsible for reviewing the weighting outputs and performing the necessary quality checks on their national data at each step of the weighting process. In Germany, the National Centre reviewed the different weighting forms delivered by the Consortium as well as the weights, and, after some minor iterations with the Consortium, approved them.

5.2 Extended Nonresponse Bias Analyses

The PIAAC standards require countries to limit the risk of nonresponse bias by striving for high response rates and applying measures to reduce bias. These measures include adaptive survey design techniques and weighting (OECD, 2022). However, bias may still affect data and estimates even after weighting. To assess whether any bias remained in the final data – particularly in the proficiency estimates – after weighting, extended NRBA were required for all participating countries with a response rate below 70 % (Krenzke et al., 2025; OECD, 2022).⁵² In Cycle 2, this applied to 29 of the 31 countries, including Germany.

By default, the international Consortium carried out the NRBA for the countries – including Germany – to ensure the use of comparable methods across countries.⁵³ The Consortium also provided an interpretation of the results, paying particular attention to the possibility of remaining bias in the data. For these analyses, countries provided the Consortium with additional data and relevant information. The extended NRBA comprised the following seven analyses:⁵⁴

52 Specifically, the 70 % cutoff refers to the background questionnaire response rate (see Section 4.5.1).

53 Alternatively, countries could also opt to perform the NRBA themselves; this was the case for nine countries.

54 Note: We report the original titles of the analyses as used by OECD (2023b, p. 17), which were applicable at the time the analyses were conducted. Krenzke et al. (2025, p. 444) modified the titles of some analyses for publication in the International Technical Report.

1. Comparison of estimates before and after weighting adjustments
2. Comparison of weighted estimates to external totals
3. Explained variation in outcomes
4. Comparison of estimates from alternative weighting adjustments
5. Analysis of variables collected during data collection
6. Level-of-effort analysis
7. Calculation of the range of potential bias⁵⁵

The following section provides a brief overview of each analysis. These are described in more detail in Krenzke et al. (2025, pp. 443–459; unless otherwise noted, this source will not be explicitly referenced further). Results for Germany are summarized, and for selected analyses – specifically analyses 2, 3, 6, and 7 – the findings are presented in more detail, supplemented by tables or figures for illustration. Most of these results and interpretations refer to an unpublished document, which the Consortium provided to the National Centre as documentation of the NRB analyses for Germany (Westat, 2024). In addition, all analyses for Germany were independently replicated by the National Centre using its own procedures and software tools. This allowed for a thorough verification and assessment of the Consortium’s results. The National Centre reached the same overall conclusions regarding potential bias in the German sample as the Consortium, indicating consistency in the findings.

5.2.1 Comparison of Estimates Before and After Weighting Adjustments

The purpose of this analysis was to assess how effectively the nonresponse weighting adjustments reduced unit nonresponse bias. To this end, for key variables, estimates from the base-weighted gross sample were compared with estimates from the respondents (net) sample, both *before* and *after* the nonresponse adjustment weights were applied. Comparing the percentage distributions of the gross sample with those of the net sample before weighting reveals potential differences attributable to nonresponse. If the estimates for the net sample after weighting are closer to those of the base-weighted gross sample, this indicates that the weighting procedure reduced bias in these variables.

For this analysis, the international Consortium relied on the standard non-response adjustment variables provided by countries, along with at least one additional variable chosen by each country. Including additional variables allowed to assess whether nonresponse weighting might also mitigate bias in variables beyond the core set of weighting variables. The variables used in the analyses for Germany are listed in Table 5.1.

⁵⁵ In the International Technical Report this analysis is now named “Measures of non-response error susceptibility that may remain” (Krenzke et al., 2025, p. 455).

Table 5.1: Variables Used for the Comparison of Estimates Before and After Weighting Adjustments in NRB Analysis 1

Variable (number of categories)	Used in weighting or additional variable for extended NRBA 1
Age (5)	Nonresponse adjustment
Gender (2)	Nonresponse adjustment
Citizenship (2)	Nonresponse adjustment
Federal state (14)	Nonresponse adjustment
Municipality size (10)	Nonresponse adjustment
Region (4)	Additional
Age (5) x gender (2)	Additional
Age (3) x citizenship (2)	Additional
Type of dwelling unit (8)	Additional

Note. NRB = nonresponse bias.

Overall, the results of the first nonresponse bias analysis indicate that the nonresponse adjustment in Germany effectively reduced nonresponse bias in the estimates of the variables listed above (Westat, 2024). The differences between the gross sample estimates and the net sample estimates *after* the nonresponse weighting adjustment were smaller than those observed *before* weighting.

Only minor differences between the gross sample and the nonresponse-weighted net sample remained in the estimates of citizenship, type of dwelling unit, and municipality size. Although statistically significant, none of these differences exceeded 1.5 percentage points, which is below the Consortium's threshold for notable differences (Krenzke et al., 2025, p. 445).

5.2.2 Comparison of Weighted Estimates to External Totals

In this second nonresponse bias analysis, the Consortium compared PIAAC estimates with data from high-quality external sources to determine whether potential bias remained in the final net sample data. For this purpose, countries provided benchmark data for selected variables from an external source of their choice, provided that they met the following requirements:

- The population covered by the external data could be aligned with the PIAAC target population.
- The variables used for comparison were measured in the same way as in PIAAC and referred to a time frame comparable to the PIAAC data collection period.

In Germany, the highest-quality external database available – the Microcensus – was used, as it met the required criteria. Specifically, Microcensus data from 2021 were used, as it was the most recent and accessible source at the time of the

NRBA.⁵⁶ In addition, sampling frame data were used to benchmark regional information.⁵⁷

To compare the final weighted German PIAAC data to the external sources, the National Centre selected variables that were available for all completed cases, including those for which only doorstep interview data were collected (i.e., with a reduced set of available variables). Table 5.2 presents the distribution of these selected variables in the final weighted PIAAC data, the Microcensus, and the sampling frame.

Table 5.2: Distribution of the Final Weighted PIAAC Data, Microcensus Data, and Sampling Frame Data, by Key Socio-demographic and Regional Variables

Variable	Value	Estimates from final weighted PIAAC data %	Estimates from Microcensus 2021 %	Estimates from sampling frame 2021 %
Age	16–25 years	15.80	15.80	
	26–35 years	19.91	19.91	
	36–45 years	19.23	19.23	
	46–55 years	21.97	21.97	
	56–65 years	23.08	23.09	
Gender	Male	50.55	50.55	
	Female	49.45	49.45	
Citizenship	German	84.31	84.81	
	Other	15.69	15.19	
Municipality size	Up to 1 999 residents	1.50		1.20
	2 000 to 4 999 residents	1.81		1.77
	5 000 to 19 999 residents	7.75		7.60
	20 000 to 49 999 residents	11.27		10.80
	50 000 to 99 999 residents; agglomeration/transfer/periphery area	9.23		8.70
	50 000 to 99 999 residents; core area	1.82		2.19
	100 000 to 499 999 residents; agglomeration/transfer/periphery area	14.97		14.57
	100 000 to 499 999 residents; core area	14.69		15.00
	Over 500 000 residents; agglomeration/transfer/periphery area	9.98		9.57
	Over 500 000 residents; core area	26.98		28.60

Note. Source: Calculations produced by Westat (2024).

⁵⁶ Source: RDC of the Federal Statistical Office and Statistical Offices of the Federal States of Germany, Microcensus 2021, <https://doi.org/10.21242/12211.2021.00.00.3.1.0>

⁵⁷ Source: Destatis: Municipal or territorial status as of June 30, 2021 (Gemeinde- bzw. Gebietsstand vom 30.06.2021), www.destatis.de

As shown in Table 5.2, the distributions of age and gender in the PIAAC data are identical to those in the 2021 Microcensus, as these two variables were explicitly used in the calibration step (see above). In addition, the analysis revealed no significant differences between the Microcensus and PIAAC estimates for German citizenship, nor between the sampling frame and PIAAC estimates for municipality size. This suggests that the calibration not only aligned the net sample with respect to the variables used in this particular weighting step, but also helped reduce bias in variables that were included only in the nonresponse weighting, not in the calibration itself (Westat, 2024).

5.2.3 Explained Variation in Outcomes

The two previous analyses focused on identifying potential bias in auxiliary variables (sociodemographic or regional). The goal of the next analysis is to compute a separate measure – the *Explained Variation in Outcomes (EVO)* – to examine potential bias in the key outcomes of PIAAC: the proficiency estimates. The EVO measure can be interpreted as a *weighted response rate* and is computed using the following formula (Krenzke et al., 2025, p. 446):

$$EVO = \frac{\sum_r w_i + R^2 \sum_{nr} w_i}{\sum_r w_i + \sum_{nr} w_i}$$

with

r = set of respondents
 nr = set of eligible nonrespondents
 w_i = base weight for case i
 R^2 = explained share of variance

The EVO measure relies on two components: the base-weighted response rate and the explained share of variance, obtained from a linear regression model with proficiency as the dependent variable and auxiliary (weighting) variables as independent variables. This approach assumes that bias in auxiliary variables may contribute to potential bias in proficiency estimates if these variables are correlated. The R^2 value indicates the strength of this relationship. For variables used in the weighting adjustments, a low R^2 suggests weak correlation and, therefore, limited effectiveness of the weighting variables in mitigating nonresponse bias. Conversely, a high R^2 indicates a stronger correlation, suggesting that the selected weighting variables were more effective in reducing nonresponse bias.⁵⁸

The EVO measure can take on values between 0 and 1. It has a value of 0 only if the response rate is 0% and there is no correlation between the auxiliary variables and proficiency ($R^2 = 0$). Conversely, it reaches a value of 1 only if all sampled persons participated in PIAAC (response rate of 100%) or if there is a perfect correlation between the auxiliary variables and proficiency ($R^2 = 1$).

⁵⁸ It should be noted, that R^2 is computed using data from respondents only. Potentially strong correlations between auxiliary variables and proficiency among nonrespondents are unknown but could be different from what is observed for respondents.

Table 5.3 presents the results for Germany, showing the R^2 values and the EVO measures for each proficiency domain. Two sets of results are provided per domain, using different variables as predictors in the regression models for proficiency: (a) variables included in the nonresponse adjustment step and (b) variables included in the nonresponse adjustment step and variables included in the calibration step. For each domain, a separate model was estimated with a different dependent variable (i.e., proficiency scores for literacy, numeracy, and adaptive problem solving). As a result, both the resulting R^2 values and the calculated EVO measures differ across the domains.

Table 5.3: Explained Share of Variance, Explained Variation in Outcome, and BQ Response Rate, by Proficiency Domain and Different Types of Weighting Variables

Proficiency domain	Variables used in weighting step	R^2	SE (R^2)	EVO	SE (EVO)	Overall BQ RR
Literacy	NR adjustment	0.332	0.0170	0.624	0.0025	45.3 %
	NR adjustment & calibration	0.528	0.0126	0.734	0.0030	45.3 %
Numeracy	NR adjustment	0.279	0.0166	0.594	0.0020	45.3 %
	NR adjustment & calibration	0.496	0.0125	0.717	0.0029	45.3 %
Adaptive problem solving	NR adjustment	0.296	0.0164	0.604	0.0035	45.3 %
	NR adjustment & calibration	0.485	0.0131	0.710	0.0031	45.3 %

Note. Source: Calculations produced by Westat (2024). The first plausible value per domain was used as the dependent variable for proficiency. Plausible values are proficiency estimates derived by the international Consortium using Item Response Theory (IRT) modeling (OECD, 2025c). BQ = background questionnaire; EVO = explained variance in outcomes; NR = nonresponse; R^2 = explained share of variance; RR = response rate; SE = standard error.

The results show that, for each of the three domains, models that included variables from both weighting steps (nonresponse adjustment and calibration) had higher R^2 values and larger EVO measures than models that used nonresponse adjustment variables alone. Although the background questionnaire response rate in Germany was 45.3 %, the weighting adjustments applied in both steps proved effective in reducing potential bias. As a result, the realised sample conveys more information than the response rate figure by itself suggests. Specifically, the explained variation is 73.4 % for literacy, 71.7 % for numeracy, and 71.0 % for adaptive problem solving. The extent to which the EVO values exceed the background questionnaire response rate indicates the relative effectiveness of the weighting adjustments in mitigating bias (Westat, 2024).

5.2.4 Comparison of Estimates From Alternative Weighting Adjustments

The purpose of this analysis was to examine whether alternative weighting variables could have achieved a greater reduction in bias than those used in the calibration step. Such analyses are particularly informative when considering variables that may be correlated with both response status and proficiency, but which were not available at the time the weighting for PIAAC was carried out. Since the proficiency measures themselves were not available during the weighting process, it was

not possible to assess potential associations with other variables at that stage. Only retrospectively, once the proficiency measures became available, was it possible to evaluate whether additional variables were closely related to proficiency (and response status). Although such alternative variables could not be used for calibration in PIAAC Cycle 2, this nonresponse bias analysis provides useful insights for future PIAAC cycles and other surveys. To assess the impact of alternative weighting adjustments on potential bias in the proficiency estimates, the Consortium compared the mean proficiency estimates for the overall adult population and for different socio-demographic groups (e.g., by age, gender, and citizenship) before and after the reweighting.

In Germany, the options for alternative weighting variables were limited, as the original weighting procedure had already included the most suitable variables available. Nonetheless, a different, more fine-grained categorization of the age variable was used and combined with the binary gender variable for reweighting: Whereas the original weighting used five age groups, the alternative approach applied ten. The final weighted data were then recalibrated to the 2021 Micro-census using this refined age-by-gender variable.

The comparison of the mean proficiency estimates weighted by the final PIAAC weights with those weighted by the alternative weights revealed no or only very small changes in the proficiency estimates. While some statistically significant differences between the proficiency estimates were observed when comparing the two weighting approaches, the magnitude was minimal (0.21 to 0.74) and therefore considered not notable (Westat, 2024).⁵⁹ Thus, the alternative weighting adjustments indicated that the selected weighting variables had already effectively reduced bias, and that using a finer grained age-by-gender variable would not have further improved bias reduction.

5.2.5 Analysis of Variables Collected During Data Collection

Disposition codes document the reasons why individuals did not participate in PIAAC. Among these, literacy-related reasons are of particular importance, as such cases are treated separately in the weighting process (see above). The present analysis examined the distribution of literacy-related nonrespondents in greater detail by comparing them with other nonrespondents to assess whether the groups differed as expected. Since literacy-related nonrespondents are assumed to have distinct characteristics (e.g., a higher proportion with migration background), the analysis was expected to reveal systematic differences between the two groups. Chi-square tests were performed to determine the relationship between the disposition codes and the selected sociodemographic variables.

Statistics from official data sources in Germany show that it is mainly individuals of prime working age who immigrate to the country, and that the proportion of individuals with a migration background or language barriers varies substantially across regions (federal states). Thus, compared to other nonrespondents, literacy-related nonrespondents were expected to include a larger propor-

59 The Consortium's definition of a notable difference is "a significant difference for which the absolute difference is greater than the standard error of the mean for the final weighted sample" (Krenzke et al., 2025, p. 449).

tion of non-German citizens, a higher share of individuals in the prime age group of 26–45 years, and reflect regional differences across federal states.

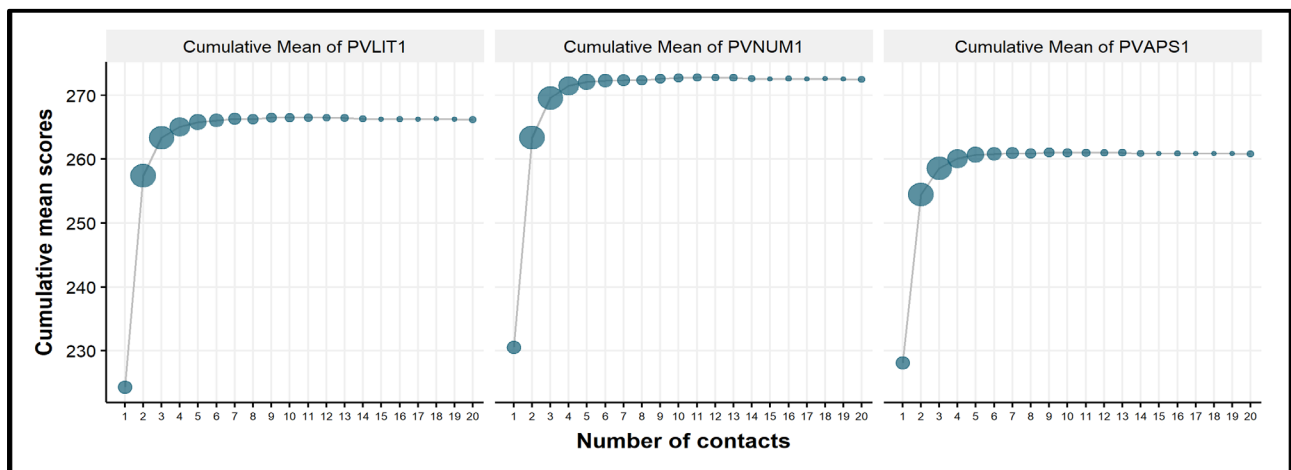
The results largely confirmed these expectations. Literacy-related nonrespondents were more likely to be non-German citizens and to fall within the 26–45 age group than other nonrespondents. Moreover, consistent with the assumptions, significant differences were observed in the distribution across municipality sizes. As no unexpected patterns emerged, the analysis provided no indication of issues with the assignment of literacy-related disposition codes or with the treatment of literacy-related nonrespondents in the weighting procedure in Germany (Westat, 2024).

5.2.6 Level-of-Effort Analysis

This analysis compares proficiency estimates for respondents who were reached with different levels of effort (Krenzke et al., 2025). The underlying assumption is that “hard-to-reach” respondents, who require more interviewer contacts, are more similar to nonrespondents, and substantial differences between “hard-to-reach” and “easy-to-reach” respondents (those contacted with fewer attempts) may indicate potential nonresponse bias in the proficiency estimates (Curtin et al., 2000).

For Germany, Figure 5.1, produced by the Consortium, shows the cumulative mean proficiency scores of respondents (y-axis) by the number of contact attempts required by interviewers to get an interview with the respondent on the x-axis (Westat, 2024).

Figure 5.1: Comparison of Cumulative Proficiency Estimates by Level of Effort



Note. Source: Westat (2024). The size of each bubble represents the relative number of respondents participating after each number of contacts. Only the first plausible value was used in this analysis to calculate the proficiency scores. PVAPS1/ PVLIT1/ PVNUM1 = variable name (in PIAAC dataset) for first plausible value for adaptive problem solving/literacy/numeracy.

Across all three domains, during the initial contact attempts – up to the third attempt – the cumulative mean proficiency scores increase steadily with each additional contact. Beyond the first few attempts, the cumulative mean scores change very little. Thus, respondents who were contacted with fewer attempts had somewhat lower mean proficiency scores than those who were harder to reach. The

differences observed between “hard-to-reach” and “easy-to-reach” respondents suggest that field efforts were successful in reaching individuals with different characteristics – potentially similar to nonrespondents – and therefore contributed to reducing potential bias in the German data (Westat, 2024).

5.2.7 Calculation of the Range of Potential Bias

As previously mentioned, one of the goals of weighting is to reduce bias arising from nonresponse, as not all sampled individuals participate in the survey. However, it is unlikely that weighting adjustments eliminate all such bias. This analysis estimates the extent of potential bias in proficiency estimates that may remain in the data after weighting.

Bias due to nonresponse can only be reduced through weighting if the available weighting variables are related to both proficiency and response status, and if certain assumptions are met. Krenzke et al. (2025) summarize the assumptions underlying the nonresponse adjustment step where adjustment cells are formed to group individuals with similar characteristics: (a) response propensities are uniform within each adjustment group/cell, (b) proficiency shows no variance within each cell, and (c) response propensity is not related to proficiency. The authors note that, in practice, these assumptions are not fully met and that information on the proficiency levels of nonrespondents is not available.

The analysis of the range of bias (RoB), as described by Krenzke et al. (2025), estimates the extent to which a hypothetical mean proficiency score – assuming full participation of all sample individuals – would deviate from the observed mean proficiency score in the PIAAC data. Given that the proficiency of nonrespondents is not known, different assumptions were applied to approximate their mean proficiency score within each weighting cell defined in the nonresponse adjustment step. Two hypothetical scenarios were considered: As a lower bound, nonrespondents were assumed to score at the 40th percentile of the respondents’ proficiency distribution within their respective weighting cell, while as an upper bound, the 60th percentile was used. The difference between these two hypothetical mean proficiency values for nonrespondents defines the potential RoB. The size of the RoB is mainly determined by two factors. First, the degree of homogeneity in proficiency scores within each weighting cell: the more homogeneous the scores within a cell, the smaller the RoB. Second, the response rate: with higher response rates, even a substantial variation in the proficiency scores within a cell has limited impact on the RoB, as nonrespondents represent only a small share of the weighting cell.

Table 5.4 shows the results for the RoB analyses for Germany, computed by the international Consortium (Westat, 2024). The RoB amounts to 14 points for literacy, 14 points for numeracy, and 12 points for adaptive problem solving. These results suggest that there is some potential for nonresponse bias in the German data. For literacy, for example, the average proficiency score would have been 13 points below the observed mean if all nonrespondents had scored at the 40th percentile. Conversely, the average proficiency score would have been 1 point higher than the observed mean if all nonrespondents had scored at the 60th percentile. However, the discussion remains hypothetical, as the underlying assumptions about nonrespondents’ proficiencies cannot be verified.

Table 5.4: Range of Potential Bias in the German Data, by Proficiency Domain

Domain	Score R, base weight	NR score 40th per- centile	NR score mean	NR score 60th per- centile	Score R, final weight	Lower bias	Higher bias	Range of bias
Literacy	272.85	265.32	270.06	279.29	266.18	-13.11	0.86	13.97
Numeracy	278.55	271.67	276.70	285.46	272.48	-12.99	0.81	13.80
Adaptive problem solving	266.06	259.49	263.91	271.33	260.81	-10.52	1.32	11.84

Note. Source: Calculations produced by Westat (2024). Only the first plausible value was used in this analysis to calculate the proficiency scores. Share of respondents = 43.8 %. R = respondents; NR = nonrespondents.

Although some of the NRBA suggested that the German PIAAC data may still be affected by potential bias, the Consortium concluded overall for all countries, including Germany, that weighting contributed to a reduction of potential bias (Krenzke et al., 2025).

5.3 Variance Estimation

In large-scale assessments such as PIAAC, the use of sampling weights is essential to ensure that the survey results provide valid estimates for the target population. Equally important are reliable variance estimators that account for the variation introduced through sample design, selection, imputation of proficiency scores, and weighting adjustments. A harmonized estimation approach minimizes effects from different weighting approaches and variance estimation procedures across countries, thereby supporting more reliable and comparable results when drawing conclusions at the international level (OECD, 2022).

Most participating countries, including Germany, implemented complex sample designs that introduced *sampling variance* through clustering. To estimate proficiency, the international Consortium applied Item Response Theory (IRT) models, treating proficiency as a latent trait, and imputed ten plausible values per domain for each respondent (OECD, 2025c). This procedure, however, introduced *imputation variance* into the data.

To account for the different sources of variation, PIAAC uses the replication approach and offers four methods: *delete-one jackknife*, *paired jackknife*, *balanced repeated replication (BRR)*, and *BRR with Fay's method* (Van de Kerckhove, Alvarez-Rojas, et al., 2025, p. 236). Replication is well suited to an international survey context because it accommodates a wide range of sampling designs (OECD, 2022). In consultation with the international Consortium, each country selected the method most appropriate for its sampling design. Replicate weights were generated as part of the weighting process: Between 15 and 80 subsamples of the full sample were formed, and each was weighted following the same procedures as for the full sample (Van de Kerckhove, Alvarez-Rojas, et al., 2025).

To compute standard errors, the first step is to produce estimates for the full sample (using final weights) and for each replicate (using the 1st, 2nd, 3rd, ... rep-

lication weight). In a second step, and considering the country's specific replication method, the standard error is calculated as the square root of the sampling variance, which is derived from the sum of the squared differences between replicate estimates and the full-sample estimate, using the formula provided, for example, by Van de Kerckhove, Alvarez-Rojas, et al. (2025, p. 236). This formula applies to the estimation of standard errors of statistics without plausible values (e.g., mean age). The estimation can be computationally intensive, depending on the number of replicates. For example, in a country with 80 replicates, the statistic must be calculated 81 times in order to apply the formula. The computation is even more complex when proficiency scores are involved, since statistics (e.g., mean literacy score) are estimated using ten plausible values. In this case, the statistic needs to be computed $81 \times 10 = 810$ times. Van de Kerckhove, Alvarez-Rojas, et al. (2025, pp. 238–239) provide the formulas required to estimate variance accounting for sampling and imputation error. To handle the complexity, statistical software with dedicated syntax or specific PIAAC tools are available to support users in producing valid statistics and standard errors (cp. Section 6.4.2).

With regard to the replication method in Cycle 2, the international Consortium recommended using Fay's method of balanced repeated replication with a perturbation factor of 30 % for the following reasons (Van de Kerckhove, Alvarez-Rojas, et al., 2025):

- Fay's BRR performs better than other replication methods in estimating variances of different quantiles (e.g., percentiles, median), which are widely reported with PIAAC data.
- Since no respondents are excluded when constructing replicate subsamples – as opposed to the jackknife methods – the issue of undefined statistics in small subgroups can be avoided.

With the exception of France, all countries, including Germany, followed the Consortium's recommendation and applied Fay's BRR (Van de Kerckhove, Alvarez-Rojas, et al., 2025). For Germany, the application of the Fay's BRR approach with 80 replicates in Cycle 2 represents a change in replication method compared to Cycle 1, where the delete-one jackknife approach – likewise with 80 replicates – was used (Zabal et al., 2014). However, the international Consortium argued that the two cycles are independent and can therefore use different approaches. All replication methods aim to approximate the same true variance, so trend analyses should not be affected (Westat, 2022, personal communication).

5.4 Design Effect

Countries implement the best possible sample design within their national context, while also considering fieldwork and cost constraints. Many PIAAC countries, including Germany, used complex multi-stage sample designs with clustering. In clustered samples, however, the selected units tend to share more similar background characteristics than those drawn through simple random sampling (SRS). As a result, estimates of population parameters from such complex designs (usually) show greater uncertainty – reflected in larger standard errors – than those based on SRS of the same sample size. In addition, unequal survey weights, which

often occur in a complex sample design, can further increase the variability of survey estimates. For more information on this topic, see, for example Ganninger (2010), Krenzke et al. (2025), or Lohr (2021).

To quantify the effect of complex sample designs on the variance of an estimator, the design effect can be computed. The design effect indicates the relative efficiency of the complex sampling design compared to an SRS design and is calculated as the ratio of the variance of the estimate obtained under the complex design to the variance of the estimate under SRS, assuming equal sample sizes (Lohr, 2021). The design effect for a statistic t can be computed using the following formula, where p denotes the applied complex sampling design:

$$Deff_t = \frac{Var(t)_p}{Var(t)_{SRS}}$$

A value greater than 1 indicates a loss of precision in the estimate under the applied sampling design compared to an SRS design, whereas a value less than 1 suggests a gain in precision relative to SRS. However, identifying an appropriate estimator for the denominator in the calculation is challenging. In principle, the survey would need to be repeated using an SRS design. However, this is not feasible in practice. Therefore, model-based estimates are commonly employed. The design effect can be approximated using the following formula:

$$\widehat{Deff}_t = n \frac{\sum_{l=1}^L w_l^2 m_l}{(\sum_{l=1}^L w_l^2 m_l)^2} * [1 + (\bar{b} - 1) * \hat{p}] = \widehat{deff}_p * \widehat{deff}_c$$

$\widehat{deff}_p$ represents the design effect arising from unequal probabilities of inclusion for individual units or groups of units. In this case, m_l denotes the number of observations in the l th weighting class, each associated with a weight of w_l (Ganninger, 2010). Since the German sample design was conceptualized as “self-weighted”, all individuals have the same design (or base) weight, which results in $\widehat{deff}_p = 1$.

$\widehat{deff}_c$ represents the design effect due to clustering, where $\bar{b}$ is the average cluster size, calculated as the net sample size divided by the number of sample points used in the complex sample design, and $\hat{p}$ is the estimate of the interclass correlation coefficient. Here, $\hat{p}$ serves as a measure of homogeneity within the clusters, that is, the similarity of the units within each cluster. A higher value of $\hat{p}$ indicates greater similarity of clustered units, leading to a negative effect on the estimate’s variance. Due to the German sample design, the cluster size – or the number of sampling units per sample point (not PSUs)⁶⁰ – was equal, resulting in a value of $\bar{b}$ that is equal to

$$b_1 = b_2 = \dots = b_l = \bar{b}.$$

60 PSU = Primary sampling unit (in the German context, these are municipalities, selected at the first sampling stage).

For Germany, the National Centre computed three design effects – one for each proficiency domain – using Stata and applying the formulas described above. To account for the ten plausible values, each calculation was conducted separately for each plausible value with each domain, and the results were then averaged. Table 5.5 summarizes the design effects for Germany.⁶¹

Table 5.5: Design Effects for the Three Proficiency Domains in Germany

Proficiency domain	Design effect
Literacy	1.66
Numeracy	1.67
Adaptive problem solving	1.64

Note. Sample size = 4 793.

The values across all three domains are similar, ranging from 1.64 for adaptive problem solving to 1.67 for numeracy. All design effect values exceed 1, which is expected due to clustering in the German sample design. Overall, these values indicate some loss of precision in the estimates compared to an SRS design.

⁶¹ Note: The international Consortium used a different approach to calculate the design effect, based on replicate weights. For Germany, the reported design effects in OECD (2025c) differ from those presented here.

6

Data Management, Data Products, and Overall Data Quality

This chapter outlines the data management tasks and processes involved in creating the final national PIAAC database within the framework of international standards and guidelines – with the aim of delivering data of the highest quality. Section 6.1 introduces the various sources that contribute to the final national dataset. In addition, the key responsibilities and roles in the national data management process *before* submission to the international Consortium are summarized. A special focus is placed on national coding activities, for example coding occupation and education. Section 6.2 describes the technical steps required to produce a user-friendly database, which are carried out at the national and the international level *after* the data submission to the international Consortium. The national data confidentiality approach underlying the definition and implementation of necessary data protection measures for the different PIAAC Germany data products is outlined in Section 6.3. The chapter concludes with an overview of the data products and analysis tools available to interested data users (Section 6.4), as well as an overall evaluation of the quality of the PIAAC Germany 2023 data based on a data adjudication process carried out at the international level (Section 6.5).

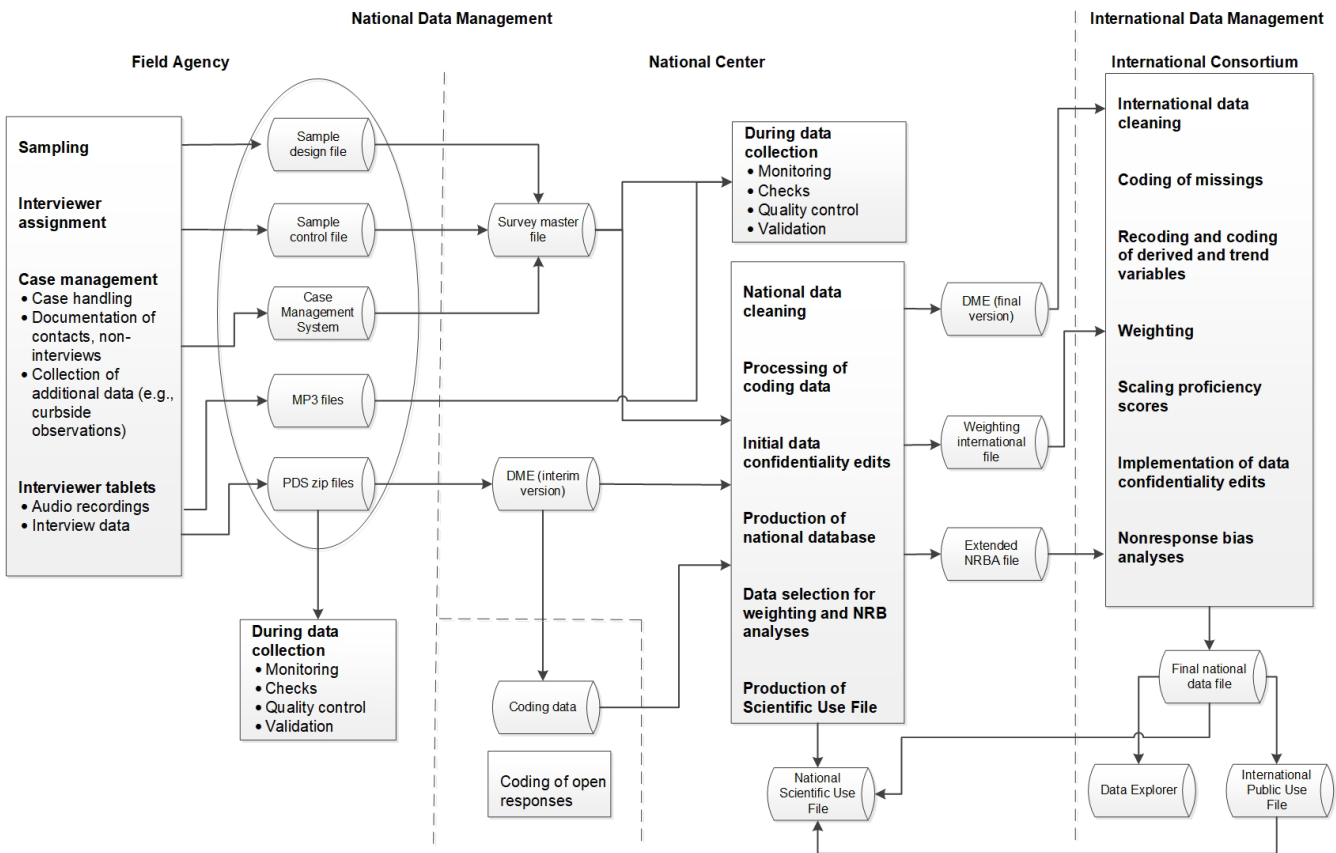
6.1 National Data Management Before Submission of National Database

National data management activities began early in the project and encompassed a wide range of tasks. Data were to be collected at various stages and from various sources, and careful planning was required for their integration. A structured data management process was established to ensure smooth processing, with clearly defined responsibilities divided between the field agency and the National Centre. The data management process had to accommodate national systems, the structures of the field agency, and international specifications.

6.1.1 National Data Management Process

Data from the PIAAC interviews followed a specific international layout and had to be imported into a customized relational database – the IEA Data Management Expert (DME) software⁶² – provided to countries by the international Consortium. Figure 6.1 illustrates the different components of the data management process, the various data sources, and how these were linked and integrated.

62 For more information on the DME, see Daniel and Gwaltney (2025).

Figure 6.1: Data Sources, Data Files, and Data Management Activities at the National and International Level

Note. The coding task was subcontracted to the International Association for the Evaluation of Educational Achievement (IEA). DME = Data Management Expert; MP3 = widespread audio format; NRB = nonresponse bias; NRBA = nonresponse bias analyses; PDS = PIAAC Delivery System.

A key starting point for the data management process before data collection was the compilation of the sampling data (sample design file). This file included essential information on each selected individual, such as personal and regional characteristics, stratification variables, and selection probabilities. The field agency was responsible for drawing the sample and created and shared the sample design file with the National Centre.⁶³

Internally, the field agency maintained and managed various databases, for example, to coordinate and monitor interviewers. They shared some of this information, such as validation results, with the National Centre (sample control file). In addition, the field agency adapted its national Case Management System (nCMS) to comply with PIAAC requirements (see Section 3.4). The nCMS contained a wide range of information, mainly entered by interviewers. This included contact (attempt) documentation, nonresponse codes, and curbside observations, for example. These data were processed by the field agency and were the basis for national monitoring activities and quality control. During fieldwork, the field agency extracted interview output files (PDS zip files) and audio recordings

⁶³ Without direct identifiers such as names and addresses.

(MP3 files) from the interviewer tablets and regularly transmitted these data to the National Centre.

Before the start of data collection, the National Centre received an initial version of the survey master file from the field agency. This file primarily contained sample design information, as well as information on interviewer assignments. During the fieldwork period, the field agency updated this file regularly (every one to two weeks) with results from the field, including data from the nCMS, and submitted it to the National Centre. This was a key resource that allowed the National Centre to obtain an overview of the fieldwork status. The data were reviewed and processed for internal monitoring and quality control purposes, as well as for international reporting.

At the National Centre, interview output files from the PIAAC delivery system were imported into the DME database, which had been customized to match the layout of the German background questionnaire. The data were regularly checked for consistency. Open-text responses – such as those relating to occupation, industry, language, and country names – were extracted on specific dates and forwarded to the subcontractor, IEA, for coding (see Section 6.1.2).

Following the data collection phase, the field agency submitted prefinal versions of both the interview and the case management data. The National Centre conducted consistency checks and cleaned the data as necessary. Final outcome codes were then assigned to all cases after reviewing the comprehensive information collected by interviewers and through a resolution process carried out together with the field agency. The next step was the production of a national database structured according to the international variable layout defined in the DME. The National Centre coded several additional variables required for the final database, such as sample design indicators, flag variables, weighting and nonresponse variables. These were then uploaded to the DME database. Basic data confidentiality measures were implemented before submitting the final national dataset to the international Consortium. These measures included removing personal identifiers and suppressing open-text entries and rare responses for less common languages and countries. Top-coding was also applied to a few selected variables. Occupation and industry codes were submitted at the 3-digit level.

A final validation check of the database was performed prior to its submission to the international Consortium. A comprehensive documentation package was delivered with the database, including details on confidentiality edits, resolution of inconsistencies, corrections to background questionnaire data, and other processing decisions. The package also contained information necessary for the weighting process, such as benchmark data for calibration (weighting international file), and for creating replicate weights for variance estimation. After the international data management activities, weighting and estimation of plausible values had been completed, the National Centre submitted a dataset with variables for the extended nonresponse bias analyses to the Consortium (extended NRBA⁶⁴ file).

64 NRBA = Nonresponse bias analyses

6.1.2 Coding of Education, Occupation, and Other Variables

It is necessary to code open text entries in order to transform them into categories that can be used in quantitative analyses. In the PIAAC background questionnaire, some questions were administered with open text entries. As a part of the national data management tasks, these open entry responses had to be coded into defined schemes according to the PIAAC international standards and guidelines (OECD, 2022). In addition, some questionnaire items were nationally adapted. For example, the questions on highest educational attainment required several questions in the German questionnaire, as opposed to only one in the international source. Recoding nationally adapted variables into the international answer scheme was also, in part, a national data management task.

Coding of Education

All questions relating to education were measured using *nationally* adapted questions and response categories. The goal was to code these adapted questions into an international coding scheme based on the *International Standard Classification of Education 2011* (ISCED 2011; UNESCO Institute for Statistics, 2012). ISCED 2011 defines educational categories using a 3-digit coding scheme. For PIAAC, a 4-digit coding scheme was developed. Official ISCED 2011 and PIAAC Cycle 1 education categories can be derived from this scheme. The scheme's enhanced differentiation allows for more detailed analyses of skills and labour market outcomes. The PIAAC coding scheme was more detailed than ISCED 2011 in several ways (for more information, see Huijts et al., 2025). The additional differentiation at ISCED 2 (lower secondary education), which allows for distinctions to be made between academically selective and general programmes, is highly relevant for Germany. Furthermore, the PIAAC scheme went into more detail at the tertiary education level, distinguishing between: (1) traditional, research-oriented universities, (2) non-traditional universities and higher education institutions, and (3) vocational qualifications at the bachelor or master level, such as master crafts qualifications. The PIAAC scheme removed differentiation at ISCED level 0 (less than primary education) because it is not relevant in most countries.

In an elaborate education consultation process between the National Centre and the Consortium (see Section 2.1.3 for more information), a list of national educational qualifications was created for each education question (highest educational attainment, current education, uncompleted education, and qualifications needed to get the current job). This process involved several iterations between the German National Centre and the international Consortium, including quality checks. For each of the education questions, the national qualifications were mapped into the international coding scheme in a correspondence table. Based on this mapping, the Consortium developed a syntax to derive international education variables. The coding scheme for mothers' and fathers' education was less detailed, with international codes limited to three levels of education (low, medium, high).

Coding of Occupation and Industry

To ensure high-quality, standardized coding procedures for occupation and industry across the participating countries, the international Consortium provided training for National Centres on recommended procedures and coding standards. In Germany, coding activities for all open entry responses (e.g., occupation, industry, languages) were carried out by the IEA in Hamburg, as part of their sub-contract with the National Centre.

As a first step prior to the coding process, the National Centre screened the data for privacy issues before transmitting the open entry data to IEA. Information that was not relevant for coding, such as company or city names, was edited and anonymized, and then checked using the four-eyes principle. Direct identifiers had already been previously removed by the field agency.

Coding was carried out by four coders from the IEA coding staff. They had several years of experience in occupation and industry coding, were native German speakers, and fluent in English (i.e., could understand the international PIAAC coding material). IEA replicated the mandatory international PIAAC coding training for the coding team. Coders were assigned 10 000 practice examples for both occupation and industry. Problematic answers were discussed in the training sessions. To pass the training, coders needed to achieve an interrater agreement of at least 85 % on all variables.

Occupation and industry were coded for the respondent's current or last job; occupation was also coded for the respondent's mother and father. All responses for occupation and industry were double-coded, exceeding the international standard of 50 % (OECD, 2022). In the event of mismatched codes, the final correct resolved code was determined by a supervisor. All results were checked for completeness and plausibility. During the coding phase, random weekly quality checks were carried out by supervisors on 10 % of the data processed by each coder.

Occupation information was coded into the 4-digit *International Standard Classification of Occupations 2008* (ISCO-08; International Labour Office, 2025). The procedure for coding occupation data was as follows:

- All data were manually coded, that is, there was no automated coding of answers according to a list of conventions.
- The coding was conducted item-wise.
- The coding sequence of answers per item was randomized to ensure that a doubtful, contingent decision did not affect several cases at once and, thus, systematically.
- In order to avoid a systematic distortion of the coding results, no simple case-of-doubt-rules (e.g., “in case of doubt assign the numerically higher code”) were employed.
- Coders were provided with some additional information (such as the industry or the economic sector of the respondent's job) to support them during the coding process.
- All data were coded into string variables. When the information provided was not sufficient to code into a substantial code, a missing code was assigned.
- When the information provided by the respondent was not sufficient to code into the 4-digit ISCO, only fewer digits, and thus less detail, were coded.

In order to comply with the PIAAC standards, the interrater reliability had to be at least 85 % at the 3-digit level, which was achieved in Germany (OECD, 2025a, p. 57). The interrater reliability ranged between 87 % and 89 % (own calculations).

Information on industry was coded into the 4-digit *International Standard Industrial Classification of All Economic Activities* (ISIC), Revision 4 (United Nations, 2008), as required by the PIAAC standards. An approach very similar to that for ISCO coding was implemented. The coding was carried out by two independent coders who double-coded 100 % of the cases. Disagreements between coders were inspected and resolved by a supervisor. The PIAAC standard, achieving an interrater reliability of at least 85 % at the 3-digit level, was also achieved for industry (OECD, 2025a, p. 57). The interrater reliability was 87 % both for industry of current as well as last job (own calculations).

After submission of the coded data by the IEA, the National Centre reviewed the data and some errors were corrected. The international Consortium reviewed and accepted the procedures and quality of the coding for ISCO and ISIC for Germany (OECD, 2025a).

Coding of Countries and Languages

Background questionnaire questions relating to languages⁶⁵ needed to be mapped or coded into the 3-digit alphanumeric code for languages, ISO 639-2/terminological (Library of Congress, 2017). Questions relating to countries⁶⁶ needed to be coded into the UN M49 numeric scheme (United Nations, 1999). In the background questionnaire, questions about languages or countries offered a closed list with the most common languages or countries of origin in Germany. Any language or country referred to by the respondent that was not included in this list was recorded as a free text entry. During the coding process, the information from the closed list was directly assigned a code by syntax. The open text information was reviewed manually and the correct code assigned. If several languages were mentioned in the open text entry, the first one mentioned was coded. Dialects were coded under the respective superordinate language (e.g., Darija – Arabic). Any typing or spelling errors that did not affect the clarity of the assignment were ignored. If a country no longer existed, the legal successor was indicated (e.g., Germany for the German Democratic Republic). A missing code (999) was assigned to typing or spelling errors that could not be clearly assigned, as well as to languages or countries that no longer exist and for which there is no clear equivalent (e.g., Serbo-Croatian). The IEA coding department carried out a quality control including checks for completeness and plausibility. After submission of the data to the National Centre, further quality checks were carried out. The National Centre also carried out research on the open text entries that had initially received a missing code and reassessed several entries; this process made it possible to assign a code in several cases.

65 First language(s) learnt and still understood; language(s) most often spoken at home.

66 Country of birth and country of the highest qualification; it was specified that the current name of the country should be indicated.

Coding of Other Information

The National Centre coded regional information derived from sampling information – namely the respondents’ place of residence (federal state) – into the TL2 level of the OECD classification of geographical regions (OECD, 2016). There were also two open questions about employment contracts in the background questionnaire that required coding (responses to “other” after a closed list). Responses were reviewed by the National Centre and, where possible, recoded into the closed list.

6.2 Data Processing After Submission of National Database

Each country submitted its national data to the international Consortium, which was responsible for managing and processing the data at the international level. To that end, the Consortium carried out several tasks to ensure that all national databases conformed to the international structure and layout. Nationally adapted variables were recoded into internationally comparable variables. As described in Chapter 2, the German background questionnaire was adapted to fit the national context. Some of the adaptations involved making structural changes, such as adding questions or response categories. During the international data processing, the Consortium recoded the national adaptations into the international PIAAC variables based on a conversion table agreed upon between the Consortium and the National Centres. In some cases, these recoding steps were highly complex, particularly the conversion of national codes for educational qualifications into the ISCED classification scheme, based on specifications defined by each National Centre (see Section 6.1.2). In Germany, several iterations of checks were required to verify that the conversion worked as intended. Aside from the standard missing values from the background interview (such as “Don’t know” or “Refused”), the Consortium also assigned additional missing codes. One example is “Valid skip”, which indicates that a question was not administered due to routing during the interview. The background data were also checked for consistency and cleaned where necessary. Any unresolved issues were addressed through iterative feedback loops with each country. In Germany, only a few issues were detected, all of which were successfully resolved. As a convenience for data users, the Consortium also generated a number of derived variables, including a comprehensive set of trend variables to simplify analyses comparing data from Cycle 2 with data from Cycle 1.

The Consortium was responsible for scaling cognitive items and estimating plausible values – imputed scores that approximate a person’s proficiency – using Item Response Theory (IRT) models. Ten plausible values were estimated for each domain. The Consortium was also responsible for conducting weighting and extended nonresponse bias analyses for several countries, including Germany. Detailed information on these activities is provided in OECD (2025c).

The production of national databases in the international format involved several rounds of review with each country. National Centres examined preliminary versions of their databases and reported any inconsistencies or errors. These issues were addressed by the Consortium, and countries received final versions of their national databases. Based on these finalized databases, the OECD and the Consortium selected a subset of variables for the Public Use Files (PUFs). To ensure maximum protection of respondent data, all PUFs underwent a thorough

data screening. Prior to the release of the PUFs, the National Centres could request the suppression and coarsening of specific variables, which the Consortium implemented and the National Centres reviewed. A separate PUF was created for each country, representing the final PIAAC data product for public users.

6.3 Data Confidentiality

As the body responsible for implementing PIAAC in Germany, the German National Centre was also responsible for ensuring compliance with applicable data protection regulations, in particular the EU General Data Protection Regulation (GDPR) and the German Federal Data Protection Act (BDSG: *Bundesdatenschutzgesetz*). All target persons selected to participate in PIAAC were sent a data protection information sheet, which informed them of all relevant data confidentiality issues, including how their data would be handled, stored, and used. It was explained that the international Consortium would receive very detailed data (but without direct identifiers), that this data would be very substantially coarsened and released to the international public via the OECD website, and that a more detailed scientific dataset would be provided to researchers who entered into a contractual agreement.

Elementary data confidentiality edits, such as removal of direct identifiers or open-entry information, was standard for all countries. Beyond that, in Germany a national data confidentiality strategy was developed that provided the framework for data confidentiality decisions for the different data products. The strategy followed the same approach as in Cycle 1 but was updated and extended for Cycle 2 (e.g., it took into account the changes in legislation brought about by the introduction of the GDPR). The basic process was identical for all data products: The National Centre reviewed every variable that was foreseen in the respective dataset (including checking univariate and in certain cases multivariate distributions), evaluated possible caveats in releasing the variable (either with or without confidentiality edits), and weighed this against its value for the specific dataset (cost-benefit analysis). After this extensive but preliminary screening, the process was reiterated with the national scientific expert for data confidentiality. At this stage, prefinal decisions were made and the basic rationale noted. As a last step, the external Data Protection Officer was given a comprehensive overview, and the key issues and decisions were reviewed and discussed again together. Adjustments were made where necessary until all parties had signed off on the decisions (unconditional release, coarsening, suppression).

Different degrees of strictness were applied in the screening and decision process. Very conservative criteria were applied to the PUF, whereas less conservative criteria were applied to the Scientific Use File (SUF), an additional national data product. The PUF is available to the general public, which comes with the highest risk. The SUF is intended to meet the much more detailed data needs of scientific data users, and at the same time requires interested researchers to sign a data use agreement in order to obtain the data. Countries did not have many degrees of freedom to select specific confidentiality edits for the PUF. The international Consortium did produce some coarsened derived variables. However, if that did not resolve country-specific issues, the only alternative was to suppress the entire vari-

able. For the SUF, the National Centre was free to implement any confidentiality edits, tailored to the German data.

6.4 Data Products, Tools, and Documentation

This section provides an overview of the available PIAAC data products. Using PIAAC data is complex due to the ten plausible values that need to be used when analyzing PIAAC skills and due to the replication weights required for variance estimation. This section also offers guidance on the tools and scripts that support the analysis of the complex PIAAC data. It also provides links to supplementary resources, such as reports, websites, code examples, and instructional videos to help users better understand the data. Although some of these resources were developed after completion of Cycle 1, they remain relevant and useful for analyzing Cycle 2 data. Most of the materials referenced in the following sections are publicly accessible without restrictions – except for the databases – and can be found on the websites of the OECD⁶⁷ and the Research Data Center PIAAC hosted by GESIS.⁶⁸

6.4.1 Data Products

The central PIAAC data products are Public Use Files (PUFs), which are subsets of the international database and include international variables only.⁶⁹ The PUFs are accessible through the OECD website.⁷⁰ The PUFs can be downloaded directly from the OECD website⁷¹ and are available to data users around the world, including government bodies, the media, and analysts.

While the database structure and variable set are consistent across countries, not all variables are populated in every national dataset. Due to country-specific data confidentiality measures, some variables may have been suppressed and therefore appear as missing in some of the datasets.⁷² The international Consortium created a set of harmonized and coarsened variables to allow some cross-country comparability despite these limitations.

In addition to the PUF, Germany has released a Scientific Use File (SUF). This is a comprehensive database that provides greater detail and enables more in-depth analyses of the German data. The SUF includes many of the variables that are suppressed or coarsened in the PUF in their original form. Moreover, it contains additional variables, nationally adapted variables, and the national extensions. The database is provided in SPSS and Stata formats.

67 <https://www.oecd.org/en/about/programmes/piaac.html>

68 <https://www.gesis.org/en/data-on-adult-education/research-data-center-piaac>

69 Specifically adapted variables or extensions at the national level are not included.

70 When this report was finalized in November 2025, PUFs were available for 30 of the 31 participating countries. The PUF for the Netherlands was not available at this time.

71 <https://www.oecd.org/en/data/datasets/piaac-2nd-cycle-database.html>

72 See document “piaac-cy2-missing_variables-puf.xlsx”, which is a list of missing variables in the PUFs and available at the OECD website.

The SUF is available exclusively to academic researchers and can be accessed through the GESIS Data Archive (study number: ZA9009, <https://doi.org/10.4232/1.14530>).⁷³ Access requires user registration and a signed data use agreement that ensures compliance with applicable data protection laws and regulations. The documentation – such as the codebook and the questionnaire – is freely accessible. Interested users can also contact the Research Data Center PIAAC at GESIS at fdz-piaac@gesis.org for further information and support.

6.4.2 Data Tools

To support data analyses, the OECD provides several tools designed to handle the complexities of the PIAAC data, including the use of replicate weights and plausible values. The following tools can be accessed on the OECD website:⁷⁴

- The *PIAAC Data Explorer*: This web-based application, operated by the OECD, produces basic analyses such as averages, percentage distributions, and frequency tables, but does not provide access to microdata or individual-level records.⁷⁵ All calculations are performed on the server side, using a database that is different from and in part more detailed than the PUFs. With its user-friendly interface, the PIAAC Data Explorer is suitable for a broad range of users – from the interested public to experienced analysts. It enables the production of valid statistical reports based on PIAAC data, accounting for plausible values and replicate weights where required. The tool includes data from all participating countries and provides an easy way to compare results from Cycle 2 and Cycle 1, which would be far more complex to achieve using the PUFs.
- The *IEA IDB Analyzer*: The International Database Analyzer is a software tool designed to support the analysis of data from large-scale assessments in education, including PIAAC.⁷⁶ Developed by the IEA, the tool is available free of charge and generates executable syntax in SAS, SPSS, or R. It enables users to perform a wide range of analyses using PIAAC data, ranging from basic statistics such as percentiles and means to more advanced methods like linear or logistic regressions. The generated code automatically accounts for the complex survey design, including the use of plausible values and replicate weights, ensuring valid statistical estimates. Analyses can be conducted on a single database such as the German SUF or a PUF. Alternatively, the *Merge Module* can be used to combine PUFs across multiple countries as needed.
- *Stata and R Modules*: The OECD has developed modules for Stata and R that allow users to perform analyses using plausible values and replicate weights. These modules are also accessible on the OECD's website. The Stata module *repest* and the R module *rrepest* are both available free of charge (Avvisati & Kessler, 2022; Ilizaliturri et al., 2025). They enable users to compute a wide range of statistics and to estimate linear or logistic regression models within the software application of their choice.

⁷³ https://search.gesis.org/research_data/ZA9009

⁷⁴ <https://www.oecd.org/en/about/programmes/piaac/piaac-data.html>

⁷⁵ <https://piaacdataexplorer.oecd.org/ide/idepiaac/>

⁷⁶ <https://www.iea.nl/data-tools/tools>

6.4.3 Documentation

All databases are accompanied by supporting documents that offer users additional guidance and context. The basic materials are codebooks, user guides, and documentation on the background questionnaire, including both the international source questionnaire and the national versions. At the international level, the OECD provides a comprehensive list of resources on its website, including numerous publications based mainly on Cycle 1 data. The following material is also available for Cycle 2:

- The *Reader's Companion* (OECD, 2024c) is a non-technical overview of the content, methodology, and implementation of the PIAAC survey. It is intended for a broad audience and helps readers to understand what the survey measures, how the results are reported, and how they should be interpreted.
- The *International Technical Report* (OECD, 2025c) is the counterpart to this National Technical Report. It presents the planning, design, and implementation of the survey from the perspective of the international Consortium and focuses on the international activities and reporting at the international level.
- The *OECD Data Analysis Manual* (OECD, 2025b) helps researchers and data users work with the PUF data. It introduces the available datasets, explains how to access them and provides information on the various data collected in PIAAC. In addition, it offers practical guidance for analyzing the PIAAC data and gives an overview of the tools offered by the OECD (as described in this section).

In Germany, the Research Data Center PIAAC at GESIS also offers additional documents and resources, including the following:

- A *collection of analysis scripts* supports open science, improves reproducibility, and guides researchers – especially early-career scholars – in analyzing PIAAC data (Maehler, 2023).
- A *digital platform* complements the previous document and provides access to the analysis scripts based on previous research with PIAAC data. These can be downloaded from <https://gesisbox.gesis.org/index.php/s/o8b2YZWqc77Nd7F>.
- An *open access methodological book* offers a practical introduction to established methods for analyzing PIAAC (Maehler & Rammstedt, 2020). Given the complexity of the data, which presents challenges for users, the book provides standardized approaches, practical examples, and software-specific tools for statistical analysis using programs such as Stata, R, or Mplus.
- A number of *online tutorials* (Research Data Center PIAAC, 2021) support users in understanding and applying PIAAC tools and software and in conducting analyses with PIAAC data (e.g., Gorges & Hollricher, 2023). These tutorials are available at <https://www.gesis.org/en/data-on-adult-education/workshops>.

6.5 Overall Data Quality

PIAAC has a well-defined program of quality assurance and quality control. Efforts are undertaken throughout the survey life cycle to achieve internationally comparable data of the highest quality possible. Quality assurance can be regarded as an upstream process which proactively works towards ensuring quality. To this end, the PIAAC Technical Standards and Guidelines (OECD, 2022) specify best practices and methodologies for all stages of the survey life cycle. These standards are binding for all countries participating in PIAAC.

Quality control can be regarded as a downstream process that tests and verifies whether standards have been met. The international Consortium has an elaborate system in place to monitor countries' compliance with PIAAC standards. A strict timeline is established, with structured input required from countries to collect key indicators and detailed information by means of assorted forms, reports, direct communication with national teams, as well as verification and sign-off processes. By closely linking quality assurance and control, it is possible to identify potential problems at an early stage and to undertake appropriate corrective action.

Data adjudication is the final and decisive quality control step in PIAAC. Its goal is to evaluate the global quality of each national dataset, its *fitness for use* (OECD, 2025a). Thus, data adjudication determines whether the quality of the PIAAC data is sufficient for its intended purpose or whether certain limitations should be imposed on the data release and cautionary notes added to its public dissemination. The adjudication framework evaluates compliance with the standards and takes data quality indicators into account.

The OECD describes the process of data adjudication in PIAAC and its results (OECD, 2025a). For each country, the Consortium collated information on *sampling, coverage and nonresponse bias, data collection, and instrumentation*. Appropriate quality indicators were defined for each of these key domains. The Consortium reviewed all the documentation and data from each country and arrived at a carefully considered evaluation for each indicator. One of three possible assessment outcomes were possible: a *pass*, when all the criteria were fully met; a *caution*, when the criteria were met to an acceptable extent; and a *fail*, when the criteria were not met sufficiently. The Consortium presented their evaluations and discussed these as well as the potential impacts on data quality with the Technical Advisory Group and the OECD. The discussion concluded with an agreement on the final outcomes which were presented to and accepted by the Board of Participating Countries.

The main results of the PIAAC adjudication report for Germany (OECD, 2025a, pp. 55–57) can be summarized as follows:

- *Sampling*: A *pass* was given for all indicators (sampling plan, sample selection, sample weighting). Regarding the sampling error: an adequate level in terms of the effective sample size was attested.
- *Coverage and nonresponse bias*: A *pass* was given for population coverage and for the basic NRBA. A *caution-low* was granted for the extended NRBA. Response rates have a pivotal role in PIAAC, and countries with a response rate below 50 % were automatically assigned a caution. However, the extended NRBA showed good results for the German data. Specifically, the indicator for the explained variation in outcomes (EVO), which can be regarded as being

similar to a weighted response rate, was above 70 % and thus warranted a *low-level* caution. To contextualize this result: Only 6 PIAAC countries received a pass (due to the strict enforcement of the response rate criterion), 13 countries were assigned a *caution-low*, 9 countries a *caution-medium*, and 2 countries a *caution-high*. In the adjudication report for Germany, it was noted that no differences were found when alternative weighting was implemented and that no differences were found to the external totals examined in the NRBA. Thus, these results did not indicate any potential for bias. The conclusion was that, given the response rate in Germany was under 50 %, “users should be cautioned that the potential for non-response bias exists in this country’s data. Relative to other countries that received a caution, there is less evidence of a potential bias” (OECD, 2025a, pp. 55–56).

- *Data collection*: A pass was assigned for all indicators (interviewer training, monitoring and validation of interviewers’ work).
- *Instrument data quality*: A pass was awarded for all indicators (translation, cognitive assessment data,⁷⁷ coding,⁷⁸ background questionnaire, item non-response).

The data adjudication report confirms that the overall quality of the German data for PIAAC Cycle 2 has the necessary quality to be included in the international dataset and reporting, that is, the results attest the fitness-for-use of the German data.

⁷⁷ This was important for PIAAC, as it indicates that the scales for all three cognitive domains are reliable, valid, and comparable (OECD, 2025a, p. 12).

⁷⁸ This refers to the reliability of the (human) coding of the open responses to occupation and industry.

Appendix A

Doorstep Interview Administration

In Germany, the *Language Identification Card* (see Figure A.1) displayed a list of 18 available foreign languages, with the sentence “Point to the language you best understand” printed in each language. This card was used during the initial interaction at the doorstep if the person answering the door did not speak (sufficient) German. The interviewer was expected to first ask whether another household member was available to assist with communication. If this was not the case, they presented the Language Identification Card to identify the language spoken by the person answering the door. By pointing to the relevant line in the list, the person was invited to indicate their preferred language for further communication.

Figure A.1: Language Identification Card

1	በተሻለ ወደሚረዱት ቋንቋ ያመለክቱ	አማርኛ	Amharisch
2	أشِر إلى اللغة التي تفهمها بالشكل الأفضل	العربية	Arabisch
3	请指明您最容易理解的语言	简体中文	Chinesisch (vereinfacht)
4	選擇您最能夠掌握的語言	繁體中文	Chinesisch (traditionell)
5	Point to the language you best understand	English	Englisch
6	اشاره كنيد به زبانی كه بيشتر و بهتر موجه می‌شويد	فارسی	Farsi
7	Indiquez la langue que vous comprenez le mieux	Français	Französisch
8	Indicare la lingua che comprende meglio	Italiano	Italienisch
9	Pokažite na jezik koji najbolje razumijete	Hrvatski	Kroatisch
10	Proszę wskazać język, który rozumie P. najlepiej	Polski	Polnisch
11	Aponte para a língua que compreende melhor	Português	Portugiesisch
12	Indicați limba pe care o înțelegeți cel mai bine	Română	Rumänisch
13	Укажите язык, который Вы понимаете лучше всего	Русский	Russisch
14	Tilmaan luuqada aad sida ugu fiican u fahmeysid	Soomaali	Somali
15	Señale el idioma que entienda mejor	Español	Spanisch
16	En iyi anladığınız dili gösterin	Türkçe	Türkisch
17	Укажіть мову, яку Ви розумієте найкраще	Українська	Ukrainisch
18	اس زبان کی طرف اشارہ کریں جس کو آپ بہت اچھی طرح سمجھتے ہوں	اردو	Urdu

Note. Adapted from International PIAAC Consortium (2020c).

After the language had been identified, the interviewer presented the *Doorstep Introduction Card* (see Figure A.2) in the previously selected language. Given the communication challenges at the doorstep, the main purpose of this card was to clarify whether the person answering the door was the target person. To this end, the Doorstep Introduction Card provided a brief introduction and explained the purpose of the interviewer's visit. It indicated that participation in the survey was entirely voluntary and that all responses would be handled with strict confidentiality and then asked whether the target person was available for a short interview. One of four possible outcomes resulted from this: (1) the contact person was the target person, (2) the target person was at home and could come to the door, (3) the target person was unavailable, or (4) the target person did not live at the address. The latter two outcomes required the interviewer to either continue with contacting attempts – possibly at a different address if the target person had moved – or to record a nonresponse. If the target person was present and able to come to the door and could speak German, the interviewer continued with the standard procedure for a regular PIAAC interview. If the target person could not speak (sufficient) German, both the Language Identification Card and the Doorstep Introduction Card were re-used. If the target person indicated that none of the languages listed on the Language Identification Card matched their preferred language, the case was documented as a *language-related nonresponse*. Language-related nonresponse was also coded if no communication with the target person was possible due to the language barrier or if the interviewer was unable to obtain any information about the language spoken by the target person. However, if basic communication between the interviewer and target person was possible, the interviewer was first required to check whether the target person was willing to participate in the full-length background interview with the support of an interpreter. The third card, the *Doorstep Consent Card* (see Figure A.3), was only presented to the target person when this option was ruled out.

Figure A.2: Doorstep Introduction Card

Doorstep Introduction Card - Registry

Hello, I am working on a study for Kantar that collects information on work and educational activities of adults.

Participation in this study is voluntary and very important. All of the answers collected will be kept strictly confidential and private.

 (SP Name) [*interviewer instruction: please point to the name of the target person on the address sheet*] has been selected to participate in the study. Is this person available now for an interview?

Please choose an answer and point to it.

1. Yes, I am this person
2. Yes, I can ask this person to come to the door
3. No, this person is unavailable
4. No, this person does not live here

Note. Adapted from International PIAAC Consortium (2020b).

The Doorstep Consent Card was presented to the target person in their chosen language and provided additional information about the interview, including that the target person would enter their responses on a tablet. The primary purpose of the card was to obtain the target person's consent to participate in the doorstep interview. In this context, it was re-iterated that participation in the doorstep interview was entirely voluntary and that all collected data would be treated with strict confidentiality. At this stage, the interviewer handed over the data protection sheet in German and, if applicable, in one of the available translated versions (Arabic, English, Polish, Romanian, Russian, or Turkish). If the language spoken by the target person was not covered by one of the translations, the English version of the data protection sheet was provided in addition to the German one.

Interviewers also handed the target person a multilingual FAQ sheet (either in English or, if available, in the appropriate foreign language version). Target persons indicated their consent by pointing to the corresponding response on the Doorstep Consent Card ("1. Yes"). If the target person pointed to "2. No", the outcome was recorded as a language-related nonresponse. If the target person agreed to participate in the doorstep interview, the interviewer launched the doorstep interview in the selected language on the tablet and then handed the device over. The target person then completed the short set of questions.

Figure A.3: Doorstep Consent Card

Doorstep Consent Card

Please answer a few short questions about your background, education, and work using this tablet.

The questions are easy and will only take a couple of minutes.

Your participation is voluntary and very important. All of your answers will be kept strictly confidential and private. All information will be reported for a group as a whole and your responses will not be linked to your name.

Do you agree to answer these few questions now?

Please choose an answer and point to it.

1. Yes
2. No

Note. Adapted from International PIAAC Consortium (2020a).

Appendix B

PIAAC 2023 Advance Letter

Advance Letter in German (Original)



##123456##

Vorname Nachname

Straße

XXXX Ort

Ort, Monat 2022

PIAAC – Die Internationale Studie zu Alltagsfähigkeiten Erwachsener

Sehr geehrter Herr Nachname,

mit diesem Schreiben laden wir Sie herzlich ein, an der PIAAC-Studie teilzunehmen. Das Ziel der Studie ist es, herauszufinden, welche grundlegenden Fähigkeiten Erwachsene in ihrem Leben erworben haben, sowohl in Deutschland als auch im internationalen Vergleich. Die Erkenntnisse liefern eine wichtige Grundlage für politische Entscheidungen und gesellschaftliche Veränderungen. In Deutschland wird PIAAC im Auftrag des Bundesministeriums für Bildung und Forschung unter der wissenschaftlichen Leitung von GESIS – Leibniz-Institut für Sozialwissenschaften durchgeführt.

Wie wurden Sie ausgewählt?

Für PIAAC werden mehrere Tausend Personen im Alter von 16 bis 65 Jahren befragt. Da nicht alle Erwachsenen befragt werden können, wurden zufällig verschiedene Personen aus den Einwohnermeldeämtern ausgewählt. Selbstverständlich ist Ihre Teilnahme freiwillig. Es ist jedoch sehr wichtig, dass alle ausgewählten Personen mitmachen, damit die Ergebnisse aussagekräftig sind. Ihre Teilnahme ist daher besonders wertvoll.

Haben Sie Fragen zum Datenschutz und zur Befragung?

Alle Beteiligten bei PIAAC haben sich der Einhaltung des Datenschutzes verpflichtet. Genauere Informationen finden Sie in der beiliegenden Datenschutzerklärung.

Weitere Informationen zu PIAAC gibt es im beigelegten Faltblatt oder im Internet unter www.gesis.org/info/piaac. Diesem Brief sind 5 € beigelegt als Dankeschön für die Zeit, die Sie sich dafür nehmen, unsere Studieninformationen anzuschauen.

Was passiert als Nächstes?

Die Befragung wird von dem Umfrageinstitut Kantar (ehemals Infratest) durchgeführt. In den nächsten Wochen wird sich Frau ABC von Kantar mit Ihnen in Verbindung setzen, um einen passenden Termin für die Befragung zu vereinbaren. Sie wird ihren Interviewerausweis vorzeigen und beantwortet gerne Ihre Fragen.

Nach Ende der Befragung erhalten Sie als Zeichen unserer Anerkennung für Ihre Teilnahme weitere 50 € in bar. Wir freuen uns, wenn Sie an dieser wichtigen Befragung teilnehmen, und bedanken uns schon im Voraus ganz herzlich für Ihre Unterstützung.

Mit freundlichen Grüßen

Prof. Dr. Beatrice Rammstedt
Projektleiterin PIAAC Deutschland

GESIS – Leibniz-Institut für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim

Günter Steinacker
Projektleiter Kantar

Kantar GmbH
Landsberger Straße 284
80687 München

www.gesis.org/info/piaac | piaac-studie@kantar.com | kostenlose Rufnummer: 0800/100 1425 (Mo-Fr: 9-17 Uhr)

Advance Letter in English (Back Translation)



##123456##

First name

Surname

Street

XXXXX City

City, Month 2022

PIAAC – The International Study on Everyday Skills of Adults

Dear Mr. Surname,

We are writing to cordially invite you to participate in the PIAAC study. Its objective is to find out which basic skills adults have acquired during their lives, both in Germany and internationally. The study results will provide an important basis for policy decisions and societal changes. In Germany, PIAAC is conducted on behalf of the Federal Ministry of Education and Research under the scientific management of GESIS – Leibniz Institute for the Social Sciences.

How were you selected?

Several thousand individuals aged between 16 and 65 will be interviewed for PIAAC. Since we cannot interview all adults, a number of individuals were randomly selected from the records at registration offices. Your participation is of course voluntary. However, it is very important that all selected persons participate to ensure that the results are meaningful and useful. Your participation is therefore especially valuable.

Do you have questions regarding data protection and about the interview?

All those involved in conducting PIAAC have vowed to comply with data protection regulations. You can find more detailed information about this in the attached data protection sheet.

You can find additional information about PIAAC in the enclosed flyer, as well as online at www.gesis.org/info/piaac. The € 5 enclosed with this letter are a thank-you for the time you are taking to look at our information about the study.

What will happen next?

The interview will be conducted by the Kantar survey institute (formerly: Infratest). In a few weeks, Ms. ABC from Kantar will contact you in order to arrange a suitable time for the interview. She will present her interviewer ID and will be happy to answer any questions you may have.

When the interview has been completed, you will receive another € 50 in cash as a token of our appreciation for your cooperation. We look forward to your participation in this important survey, and we would like to thank you very much in advance for your support.

Kind regards,

Prof. Dr. Beatrice Rammstedt
PIAAC Deutschland project lead

GESIS – Leibniz-Institut für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim

Günter Steinacker
Kantar project lead

Kantar GmbH
Landsberger Strasse 284
80687 Munich

Appendix C

Post-Interview Questions

International Post-Interview Questions – in German (Translation) and English (Source)

War außer Ihnen und der befragten Person noch eine weitere Person während des Interviews anwesend? Bitte wählen Sie die Antwort(en) aus, die am besten passt bzw. passen. [Bitte wählen Sie alles Zutreffende aus.] 1. Es war keine weitere Person anwesend. 2. Während der Beantwortung des Hintergrundfragebogens war jemand anwesend. 3. Während der Bearbeitung des Aufgabenteils war jemand anwesend.	Please select the option or options below that best describe(s) the presence of a person other than you or the respondent during the interview. [Select all that apply.] 1. No one else was present 2. Someone else was present during the background questionnaire 3. Someone else was present during the Exercise
Hat die befragte Person während des Interviews Hilfe von jemandem erhalten? Bitte wählen Sie die Antwort(en) aus, die am besten passt bzw. passen. [Bitte wählen Sie alles Zutreffende aus.] 1. Es hat niemand der befragten Person geholfen. 2. Während der Beantwortung des Hintergrundfragebogens hat die befragte Person Übersetzungshilfen erhalten, z.B. von einer Dolmetscherin bzw. einem Dolmetscher oder einem Haushaltsmitglied. 3. Während der Bearbeitung des Aufgabenteils hat jemand der befragten Person geholfen.	Please select the option or options below that best describe(s) whether the respondent received assistance from anyone in answering any of the following? [Select all that apply.] 1. No one assisted the respondent 2. A translator or interpreter helped the respondent with the background questionnaire 3. Someone assisted the respondent during the Exercise
Was ist insgesamt Ihr Eindruck: Hat die befragte Person die Fragen des Hintergrundfragebogens verstanden? 1. Nie 2. Fast nie 3. Hin und wieder 4. Oft 5. Sehr oft	Overall, did you feel that the respondent understood the questions in the background questionnaire? 1. Never 2. Almost never 3. Now and then 4. Often 5. Very often
Haben Sie der befragten Person während der selbstständigen Bearbeitung der Aufgaben geholfen? 1. Ja 2. Nein	Did you assist the respondent with the Exercise when the respondent was working independently? 1. Yes 2. No

Hat die befragte Person bei der selbstständigen Bearbeitung der Aufgaben um Hilfe im Umgang mit dem Tablet gebeten (hat sie z.B. gefragt, wo eine Antwort eingegeben werden soll)? 1. Ja 2. Nein	Did the respondent ask for assistance in using the features of the tablet (e.g. asked where to enter a response on the tablet) when working independently on the Exercise? 1. Yes 2. No
Hat die befragte Person bei der Bearbeitung der Aufgaben um Hilfe bei irgendeiner Frage gebeten? [Nachfragen zum Umgang mit dem Tablet sind hier nicht gemeint.] 1. Ja 2. Nein	Did the respondent ask for assistance on any questions while completing the Exercise (please exclude questions related to using the tablet features)? 1. Yes 2. No
Gab es während der Bearbeitung des Aufgabenteils Hintergrundgeräusche von einer der folgenden Quellen? [Bitte wählen Sie alles Zutreffende aus.] 1. Flugverkehr, Bahnverkehr, Straßenverkehr, einschließlich Einsatzfahrzeuge 2. Laute Geräusche außerhalb der Wohnung, z.B. Bauarbeiten in der Umgebung, Rasenmäher 3. Haushaltsgeräte in Gebrauch, z.B. Staubsauger, Waschmaschine/Trockner, Bohrmaschine 4. Laufende Fernseher, Radios, Spielkonsolen oder Musik 5. Laut sprechende Personen, weinende Babys oder Kinder, Haustiere 6. Sonstige 7. Nichts davon zutreffend	During the Exercise, was there background (ambient) noise from any of the following sources? [Select all that apply.] 1. Air traffic, rail traffic, road traffic including emergency vehicles 2. Loud noises outside the house, e.g. construction in the area, lawn mowers 3. Household appliances in use e.g., vacuum cleaner, washing machine/dryer, power tools 4. A television, radio, game console or music player in use 5. People talking loudly, crying babies or children, pets 6. Other 7. Not applicable
Hat die befragte Person zu irgendeinem Zeitpunkt die Bearbeitung des Aufgabenteils unterbrochen? 1. Ja, mehrmals 2. Ja, einmal 3. Nein	Did the respondent stop working on the Exercise at any point? 1. Yes, more than once 2. Yes, once 3. No
Hat sich die befragte Person darüber beschwert, dass der Aufgabenteil zu lange ist oder zu viel Zeit in Anspruch nimmt? 1. Ja 2. Nein	Did the respondent complain that the Exercise was taking too long or taking too much time? 1. Yes 2. No
An welchem Ort wurde der Aufgabenteil bearbeitet? 1. Zu Hause (inkl. Balkon, Terrasse, Garten usw.) 2. Am Arbeitsplatz oder im Büro 3. Außerhalb der Wohnung, z.B. in einer Volkshochschule oder Bücherei	Where did the Exercise mainly take place? 1. In the home (include patio, garden, balcony, etc.) 2. Place of work or office 3. Outside the home, e.g., public library, coffee shop

Note. International source taken from International PIAAC Consortium (2022).

National Post-Interview Questions – in German (Source) and English (Back Translation)

Hat die befragte Person bei der Bearbeitung der Aufgaben einen Taschenrechner verwendet? 1. Ja, den Online-Taschenrechner 2. Ja, den GENIE-Taschenrechner 3. Ja, sowohl den Online-Taschenrechner als auch den GENIE-Taschenrechner 4. Nein, keinen Taschenrechner 5. Ich konnte es nicht sehen 6. Trifft nicht zu (Interview wurde abgebrochen)	Did the respondent use a calculator while completing the Exercise? 1. Yes, the online calculator 2. Yes, the GENIE calculator 3. Yes, both the online calculator and the GENIE calculator 4. No, no calculator 5. I could not see it 6. Not applicable (interview break-off)
Womit hat die befragte Person die Aufgaben bearbeitet? 1. Nur oder überwiegend mit einem Eingabestift 2. Nur oder überwiegend mit dem Finger 3. Eingabestift und Finger wurden etwa gleich viel verwendet 4. Trifft nicht zu (Interview wurde abgebrochen)	What did the respondent use to complete the Exercise? 1. Only or mainly a stylus 2. Only or mainly the finger 3. Used stylus and finger almost equally 4. Not applicable (interview break-off)
Hatte die befragte Person im Aufgabenteil mit den Funktionalitäten der PIAAC-Software Probleme? 1. Ja 2. Nein 3. Ich konnte es nicht sehen 4. Trifft nicht zu (Interview wurde abgebrochen)	Did the respondent have problems with the functionalities of the PIAAC software during the Exercise? 1. Yes 2. No 3. I could not see it 4. Not applicable (interview break-off)
Bitte beschreiben Sie die Probleme mit den Funktionalitäten: _____	Please describe the problems with the functionalities: _____
Im Aufgabenteil: Hatten Sie insgesamt den Eindruck, dass die befragte Person die Aufgaben gewissenhaft bearbeitet hat? 1. Ja, (fast immer) 2. Ja, häufig 3. Ja, aber eher selten 4. Nein 5. Trifft nicht zu (Interview wurde abgebrochen)	During the Exercise: Did you feel that the respondent worked conscientiously on the tasks? 1. Yes, (almost always) 2. Yes, often 3. Yes, but only rarely 4. No 5. Not applicable (interview break-off)
Im Aufgabenteil: Hat sich die befragte Person durch die Benutzung eines Smartphones oder eines anderen elektronischen Geräts ablenken lassen? 1. Ja, einmal 2. Ja, mehrmals 3. Nein 4. Trifft nicht zu (Interview wurde abgebrochen)	During the Exercise: Was the respondent distracted by the use of a smartphone or another electronic device? 1. Yes, once 2. Yes, several times 3. No 4. Not applicable (interview break-off)

Falls Sie weitere Anmerkungen zum Interview haben, schreiben Sie diese bitte möglichst detailliert hier auf:	If you have additional comments on the interview, please write them in the space below:
1. Nein, keine Anmerkung	1. No, no comments
Falls Sie Anmerkungen zu den Hygienemaßnahmen in der Interviewsituation haben (z.B. Interviewer*in und Befragungsperson haben Maske getragen, Verständigungsschwierigkeiten wegen Maske(n), Spiegelung des Tablet-Bildschirms beim Interview im Freien), bitte hier beschreiben:	If you have comments on the hygiene measures during the interview (e.g., interviewer and respondent wore a mask, communication difficulties due to the mask(s), reflection of the tablet screen during an outdoor interview), please describe them here:
1. Nein, keine Anmerkung	1. No, no comments

References

- Ali, U., & Robin, F. (2025). The PIAAC Cycle 2 assessment design. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 19–39). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Avvisati, F., & Keslair, F. (2022). *Repest: Stata module to run estimations with weighted replicate samples and plausible values* [Computer program]. Boston College Department of Economics. https://gitlab.algobank.oecd.org/edu_data/repest
- Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). Response rates in the European Social Survey: Increasing, decreasing, or a matter of fieldwork efforts? *Survey Methods: Insights from the Field*. <https://doi.org/10.13094/SMIF-2018-00003>
- Blohm, M., & Koch, A. (2021). Monetary incentives in large-scale face-to-face surveys: Evidence from a series of experiments. *International Journal of Public Opinion Research*, 33(3), 690–702. <https://doi.org/10.1093/ijpor/edab007>
- Brick, J. M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The Annals of the American Academy of Political and Social Science*, 645(1), 36–59. <https://doi.org/10.1177/0002716212456834>
- Cox, L. H. (1987). A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association*, 82(398), 520–524. <https://doi.org/10.1080/01621459.1987.10478456>
- Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, 64(4), 413–428. <https://doi.org/10.1086/318638>
- Daniel, T., & Gwaltney, D. (2025). Data management and data-cleaning procedures. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 464–487). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Das Big Five Inventar 2: Validierung eines Persönlichkeitsinventars zur Erfassung von 5 Persönlichkeitsdomänen und 15 Facetten. *Diagnostica*, 65(3), 121–132. <https://doi.org/10.1026/0012-1924/a000218>
- European Social Survey. (2019). *ESS Round 9 (2018/2019) technical report*. ESS ERIC. https://stessrelpubprodwe.blob.core.windows.net/data/round9/survey/ESS9_data_documentation_report_e03_1.pdf
- Gal, I., Alatorre, S., Close, S., Evans, J., Johansen, L., Maguire, T., Manly, M., & Tout, D. (2009). *PIAAC numeracy: A conceptual framework* (OECD Education Working Papers, Issue 35). OECD Publishing. <https://doi.org/10.1787/220337421165>
- Ganninger, M. (2010). *Design effects: Model-based versus design-based approach* (GESIS Schriftenreihe, Issue 3). GESIS – Leibniz Institute for the Social Sciences. <https://doi.org/10.21241/ssoar.26124>
- GESIS Datenarchiv für Sozialwissenschaften. (2018). *ALLBUS 2018: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. Fragebogendokumentation. Material zu den Datensätzen der Studiennummern ZA5270 und ZA5271*. GESIS – Leibniz-Institut für Sozialwissenschaften.
- Gorges, J., & Hollricher, L. (2023). *PIAAC data analyses: An introduction using the R EdSurvey package* [Online tutorial]. GESIS – Leibniz Institute for the Social Sciences. <https://www.gesis.org/data-on-adult-education/workshops/piaac-data-analyses-examples-using-r>
- Greiff, S., Graesser, A., Iliescu, D., Rouet, J.-F., Scheiter, K., Scherer, R., Gottschling, J., & Dörendal, J. (2021). PIAAC Cycle 2 assessment framework: Adaptive problem solving. In OECD (Ed.), *The assessment frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies* (pp. 155–203). OECD Publishing. <https://doi.org/10.1787/3a14db8b-en>
- Häder, S. (2015). *Stichproben in der Praxis* (GESIS Survey Guidelines). GESIS – Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/sdm-sg_014

- Harkness, J. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. Van de Vijver, & P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). John Wiley & Sons.
- Hoffmeyer-Zlotnik, J. H. P., & Wolf, C. (Eds.). (2003). *Advances in cross-national comparison. A European working book for demographic and socio-economic variables*. Springer. <https://doi.org/10.1007/978-1-4419-9186-7>
- Huijts, T., Jacobs, B., Meijer, R., Bluemke, M., Lechner, C., Massing, N., Roth, M., Schneider, S., Singh, R., Kandathil, M., Kapur, L., Ali, U., & Robin, F. (2025). Derived variables and validation of background questionnaire data. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 297–329). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Ilizaliturri, R., Avvisati, F., & Keslair, F. (2025). *Rrepest: An analyzer of international large scale assessments in education* (Version 1.5.4) [Computer program]. The Comprehensive R Archive Network. <https://doi.org/10.32614/CRAN.package.Rrepest>
- infas Institut für angewandte Sozialwissenschaft. (2016). *Interviewerhandbuch: Deutschland in Europa – Der European Social Survey (Welle 8)*. infas Institut für angewandte Sozialwissenschaft. https://stessrelpubprodwe.blob.core.windows.net/data/round8/fieldwork/germany/ESS8_fieldwork_and_interviewer_instruction_DE.pdf
- International Labour Office. (2025). *International Standard Classification of Occupations: Structure, group definitions and correspondence tables – ISCO-08*. <https://isco-ilo.netlify.app/en/isco-08/>
- International PIAAC Consortium. (2020a). *English_PIAAC_CY2_Doorstep_Consent.docx* [Unpublished document].
- International PIAAC Consortium. (2020b). *English_PIAAC_CY2_Doorstep_IntroRegistry_v2.docx* [Unpublished document].
- International PIAAC Consortium. (2020c). *English_PIAAC_CY2_Doorstep_LangCard.docx* [Unpublished document].
- International PIAAC Consortium. (2022). *PIAAC_CY2(2022_02)MS_Post-Interview_Questions.xlsx* [Unpublished document].
- Jabkowski, P., & Cichocki, P. (2025). Survey response rates in European comparative surveys: A 20-year decline irrespective of sampling frames or survey modes. *Quality & Quantity*, 59, 635–655. <https://doi.org/10.1007/s11135-024-01993-9>
- Jacobs, B., Huijts, T., Behr, D., Bluemke, M., Krämer, F., Lechner, C., Massing, N., Ortmanns, V., Schneider, S., & Zabal, A. (2025). The development of the PIAAC Cycle 2 background questionnaires. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 133–187). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Jones, S., Gabrielsen, E., Hagston, J., Linnakylä, P., Megherbi, H., Sabatini, J., Tröster, M., & Vidal-Abarca, E. (2009). *PIAAC literacy: A conceptual framework* (OECD Education Working Papers, Issue 34). OECD Publishing. <https://doi.org/10.1787/220348414075>
- Kantar Public. (2024). *PIAAC-Hauptstudie 2022/2023 Feldbericht* [Unpublished document].
- Kapidzic, S., Gauly, B., Greiff, S., Martin, S., Massing, N., Zabal, A., & Rammstedt, B. (2025). *Adaptives Problemlösen Erwachsener im internationalen Vergleich: Ergebnisse von PIAAC 2023*. Waxmann. <https://doi.org/10.31244/9783818850784>
- Krenzke, T., Lopdell, J., Van de Kerckhove, W., & Mohadjer, L. (2025). Indicators of the quality of the sample data. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 428–463). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Lechner, C. M., Anger, S., & Rammstedt, B. (2019). Socio-emotional skills in education and beyond: Recent evidence and future research avenues. In R. Becker (Ed.), *Research handbook on the sociology of education* (pp. 427–453). Edward Elgar Publishing. <https://doi.org/10.4337/9781788110426>
- Lennon, M. L., & Halderman, L. (2025). The development of the PIAAC cognitive instruments. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 64–97). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Library of Congress. (2017). *Codes for the representation of names of languages: Codes arranged alphabetically by alpha-3/ISO 639-2 Code*. Language Coding Agency. https://www.loc.gov/standards/iso639-2/php/code_list.php

- Lohr, S. L. (2021). *Sampling: Design and analysis* (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9780429298899>
- Maehler, D. B. (2023). *Analysis scripts in large-scale assessments in education*. GESIS – Leibniz Institute for the Social Sciences. <https://doi.org/10.21241/ssoar.85846.v2>
- Maehler, D. B., & Rammstedt, B. (2020). *Large-scale cognitive assessment: Analyzing PIAAC data*. Springer. <https://doi.org/10.1007/978-3-030-47515-4>
- Martin, S., Helmschrott, S., & Rammstedt, B. (2014). The use of respondent incentives in PIAAC: The field test experiment in Germany. *methods, data, analyses* 8(2), 223–242. <https://doi.org/10.12758/mda.2014.009>
- Martin, S., & Zabal, A. (2020). Converting nonrespondents in PIAAC Germany 2012 using responsive measures. *Survey Methods: Insights from the Field*. <https://doi.org/10.13094/SMIF-2020-00009>
- Martin, S., Zabal, A., Kapidzic, S., & Lenzner, T. (2022). Optimizing advance letters: Findings from a cognitive pretest. *Survey Methods: Insights from the Field*. <https://doi.org/10.13094/SMIF-2022-00008>
- Murray, T. S., Clermont, Y., & Binkley, M. (2005). *Measuring adult literacy and life skills: New frameworks for assessment* [Catalogue no. 89-552-MIE, no. 13]. Statistics Canada. <https://www150.statcan.gc.ca/n1/pub/89-552-m/89-552-m2005013-eng.pdf>
- OECD. (2013). *Technical report of the Survey of Adult Skills (PIAAC)* (3rd ed.). OECD Publishing.
- OECD. (2016). *OECD regions at a glance 2016*. OECD Publishing. https://doi.org/10.1787/reg_glance-2016-en
- OECD. (2018). *PIAAC Cycle 2 BQ draft conceptual framework*. OECD Publishing. [https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/piaac/background-questionnaire/cycle-2/PIAAC_CY2\(2018_11\)BQ_Draft_Conceptual_Framework.pdf](https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/piaac/background-questionnaire/cycle-2/PIAAC_CY2(2018_11)BQ_Draft_Conceptual_Framework.pdf)
- OECD. (2019a). *PIAAC Cycle 2 – General translation guidelines for the background questionnaire*. *PIAAC_CY2(2019_03)BQ_Translation_Guidelines_V2.pdf* [Unpublished document].
- OECD. (2019b). *PIAAC Cycle 2: Translation and adaptation guidelines for the cognitive instruments*. *PIAAC_CY2(2019_03)_Transl_Adapt_Guidelines_COG.pdf* [Unpublished document].
- OECD. (2021). *The assessment frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies*. OECD Publishing. <https://doi.org/10.1787/4b-c2342d-en>
- OECD. (2022). *Cycle 2 PIAAC technical standards and guidelines*. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/piaac/technical-standards-and-guidelines/cycle-2/PIAAC_CY2_Technical_Standards_and_Guidelines.pdf
- OECD. (2023a). *PIAAC Cycle 2 Sampling Plan Part II: PIAAC Weighting and variance estimation plans*. *PIAAC(2023_05)Sampling Plan Part II Weighting.docx* [Unpublished document].
- OECD. (2023b). *PIAAC Cycle 2 Sampling Plan Part III: Response rates and nonresponse bias analysis plans*. *PIAAC(2023_01)Sampling Plan Part III NRBA.docx* [Unpublished document].
- OECD. (2024a). *Do adults have the skills they need to thrive in a changing world? Survey of Adult Skills 2023*. OECD Publishing. <https://doi.org/10.1787/b263dc5d-en>
- OECD. (2024b). *PIAAC Cycle 2 released cognitive items*. https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/piaac/released-items/cycle-2/PIAAC_CY2_Released_Cognitive_Items_EN.pdf
- OECD. (2024c). *Survey of Adult Skills 2023 reader's companion*. OECD Publishing. <https://doi.org/10.1787/3639d1e2-en>
- OECD. (2025a). Annex G. Data adjudication in PIAAC. In OECD (Ed.), *Survey of Adult Skills 2023 technical report*. OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- OECD. (2025b). *Survey of Adult Skills 2023 data analysis manual*. OECD Publishing. <https://doi.org/10.1787/25a87a9d-en>

- OECD. (2025c). *Survey of Adult Skills 2023 technical report*. OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- OECD & Statistics Canada. (2000). *Literacy in the information age: Final report of the International Adult Literacy Survey*. OECD Publishing. <https://doi.org/10.1787/9789264181762-en>
- OECD & Statistics Canada. (2005). *Learning a living: First results of the Adult Literacy and Life Skills Survey*. OECD Publishing. <https://doi.org/10.1787/9789264010390-en>
- OECD & Statistics Canada. (2011). *Literacy for life: Further results from the Adult Literacy and Life Skills Survey*. OECD Publishing. <https://doi.org/10.1787/9789264091269-en>
- Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment*, 36(1), 149–161. <https://doi.org/10.1027/1015-5759/a000481>
- Rammstedt, B., Gauly, B., Kapidzic, S., Maehler, D. B., Martin, S., Massing, N., Schneider, S. L., & Zabal, A. (2024a). *PIAAC 2023. Grundlegende Kompetenzen Erwachsener im internationalen Vergleich*. Waxmann. <https://doi.org/10.31244/9783830999652>
- Rammstedt, B., Gauly, B., Kapidzic, S., Maehler, D. B., Martin, S., Massing, N., Schneider, S. L., & Zabal, A. (2024b). *PIAAC 2023 Online-Anhang*. Waxmann. <https://www.waxmann.com/fileadmin/media/zusatztexte/4965Anhang.pdf>
- Rammstedt, B., Roemer, L., Behr, D., Bluemke, M., Lechner, C., Dept, S., Wäyrynen, L., Soto, C., & John, O. P. (2025). Going global: 39 language versions of the BFI-2-XS. *Measurement Instruments for the Social Sciences*, 7. <https://doi.org/10.5964/miss.14067>
- Research Data Center PIAAC. (2021). *PIAAC data analysis in Stata: A practical guide* [Online tutorials]. GESIS – Leibniz Institute for the Social Sciences. <https://www.gesis.org/data-on-adult-education/workshops/online-tutorial-piaac-data-analysis-in-stata>
- Rouet, J.-F., Britt, M. A., Gabrielsen, E., Kaakinen, J., Richter, T., & Lennon, M. (2021). PIAAC Cycle 2 assessment framework: Literacy. In OECD (Ed.), *The assessment frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies* (pp. 39–64). OECD Publishing. <https://doi.org/10.1787/7b3bf33b-en>
- Sabatini, J. P., & Bruce, K. M. (2009). *PIAAC reading component: A conceptual framework* (OECD Education Working Paper, Issue 33). OECD Publishing. <https://doi.org/10.1787/220367414132>
- Schneider, S. (2025). *The hydra of comparative surveys: Measurement and harmonization of education in PIAAC Cycle 2*. SocArXiv. https://doi.org/10.31235/osf.io/cgkdu_v2
- Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS, *Journal of Research in Personality*, 68, 69–81. <https://doi.org/10.1016/j.jrp.2017.02.004>
- Statistisches Bundesamt. (Ed.). (2004). *Demographische Standards, Ausgabe 2004*. Statistisches Bundesamt. https://www.destatis.de/DE/Methoden/Demografische-Regionale-Standards/Downloads/demografische-standards-2004.pdf?__blob=publicationFile&v=2
- Statistisches Bundesamt. (Ed.). (2016). *Demographische Standards, Ausgabe 2016. Statistik und Wissenschaft Band 17*. Statistisches Bundesamt. https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEMonografie_derivate_00001549/Band17_DemographischeStandards1030817169004.pdf
- Statistisches Bundesamt. (2021). *Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2020*. Statistisches Bundesamt. https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00136608
- The American Association for Public Opinion Research. (2023). *Standard definitions: Final dispositions of case codes and outcome rates for surveys*. AAPOR. <https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf>
- Thornton, N., Krenzke, T., & Mohadjer, L. (2025). Quality control monitoring activities. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 419–427). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>

- Tout, D., Demonty, I., Diez-Palomar, J., Geiger, V., Hoogland, K., & Maguire, T. (2021). PIAAC Cycle 2 assessment framework: Numeracy. In OECD (Ed.), *The assessment frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies* (pp. 65–154). OECD Publishing. <https://doi.org/10.1787/c4221062-en>
- UNESCO Institute for Statistics. (2012). *International Standard Classification of Education: ISCED 2011*. UNESCO Institute for Statistics. <https://unesdoc.unesco.org/ark:/48223/pf0000219109>
- UNESCO Institute for Statistics. (2025). *ISCED mappings and diagrams*. <https://www.uis.unesco.org/en/methods-and-tools/isced/mapping-and-diagrams?hub=12>
- United Nations. (1999). *Standard country or area codes for statistical use*. United Nations. [https://unstats.un.org/unsd/publication/SeriesM/Series_M49_Rev4\(1999\)_en.pdf](https://unstats.un.org/unsd/publication/SeriesM/Series_M49_Rev4(1999)_en.pdf)
- United Nations. (2008). *International Standard Industrial Classification of all economic activities (ISIC): Revision 4* (Statistical Papers Series M, Issue 4/Rev.4). Department of Economic and Social Affairs. Statistics Division. https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf
- Van de Kerckhove, W., Alvarez-Rojas, L., Krenzke, T., & Mohadjer, L. (2025). Survey weighting and variance estimation. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 214–240). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Van de Kerckhove, W., Li, L., Krenzke, T., & Mohadjer, L. (2025). Sampling design. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 40–63). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- Wasmer, M., Blohm, M., Walter, J., Jutz, R., & Scholz, E. (2017). *Konzeption und Durchführung der “Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften” (ALLBUS) 2014* (GESIS Papers, Issue 2017/20). GESIS – Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.53370>
- Wayrynen, L., Behr, D., & Zabal, A. (2025). Translation, adaptation and verification of test and survey materials. In OECD (Ed.), *Survey of Adult Skills 2023 technical report* (pp. 188–213). OECD Publishing. <https://doi.org/10.1787/80d9f692-en>
- West, B. T. (2013). An examination of the quality and utility of interviewer observations in the National Survey of Family Growth. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(1), 211–225. <https://doi.org/10.1111/j.1467-985X.2012.01038.x>
- Westat. (2024). *CY2_Extended_NRBA_Analyses_(with scores)_DEU.xlsx* [Unpublished document].
- Wolf, C., Schneider, S. L., Behr, D., & Joye, D. (2016). Harmonizing survey questions between cultures and over time. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu (Eds.), *The SAGE handbook of survey methodology* (pp. 500–522). SAGE Publications. <https://doi.org/10.4135/9781473957893.n33>
- Zabal, A., & Behr, D. (2018). Anwendungsorientierte Einführung in die Übersetzung und Adaptation von Messinstrumenten. In D. B. Maehler, A. Shajek, & H. U. Brinkmann (Eds.), *Diagnostik bei Migrantinnen und Migranten: Ein Handbuch* (pp. 32–46). Hogrefe.
- Zabal, A., Martin, S., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I., & Rammstedt, B. (2014). *PIAAC Germany 2012: Technical report*. Waxmann. <https://www.gesis.org/fileadmin/piaac/Downloadbereich/TechnicalReport-ebook.pdf>

List of Figures

Figure 2.1: Overall Flow of the PIAAC Cycle 2 Background Questionnaire	22
Figure 2.2: Assessment Design	37
Figure 2.3: Path Decision Rules	38
Figure 4.1: Simple Fieldwork Response Rate and Fieldwork Outcomes (Cumulated Over Time)	95
Figure 5.1: Comparison of Cumulative Proficiency Estimates by Level of Effort	119
Figure 6.1: Data Sources, Data Files, and Data Management Activities at the National and International Level	126
Figure A.1: Language Identification Card	139
Figure A.2: Doorstep Introduction Card	141
Figure A.3: Doorstep Consent Card	142

List of Tables

Table 2.1:	Number of Linking and New Items in PIAAC 2023	34
Table 3.1:	PIAAC Interviewer Characteristics	60
Table 4.1:	Waves and Number of Cases of Main Working Phase	74
Table 4.2:	Waves and Number of Cases of Reissue Phase	75
Table 4.3:	International PIAAC Disposition Codes and National Disposition Codes for Case Initialisation	78
Table 4.4:	Overview of Interview Validation per Phone and Postal Mail.	88
Table 4.5:	Percentage of Validated Cases per Interviewer.	93
Table 4.6:	Validation Results for Different Groups	94
Table 4.7:	Final PIAAC Disposition Codes for Germany	99
Table 4.8:	Responses to Post-Interview Questions	103
Table 5.1:	Variables Used for the Comparison of Estimates Before and After Weighting Adjustments in NRB Analysis 1.	114
Table 5.2:	Distribution of the Final Weighted PIAAC Data, Microcensus Data, and Sampling Frame Data, by Key Sociodemographic and Regional Variables.	115
Table 5.3:	Explained Share of Variance, Explained Variation in Outcome, and BQ Response Rate, by Proficiency Domain and Different Types of Weighting Variables	117
Table 5.4:	Range of Potential Bias in the German Data, by Proficiency Domain	121
Table 5.5:	Design Effects for the Three Proficiency Domains in Germany	124

List of Abbreviations

AAPOR	The American Association for Public Opinion Research
ALL	Adult Literacy and Life Skills Survey
ALLBUS	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften [German General Social Survey]
API	Application Programming Interface
APS	Adaptive problem solving
BDSG	Bundesdatenschutzgesetz [German Federal Data Protection Act]
BFI	Big Five Inventory
BPC	Board of Participating Countries
BQ	Background questionnaire
BRR	Balanced repeated replication
CAPI	Computer-assisted personal interview
CMS	Case Management System
DIPF	DIPF Leibniz Institute for Research and Information in Education
DME	Data Management Expert
ESS	European Social Survey
ETS	Educational Testing Service
EVO	Explained variation in outcomes
FAQ	Frequently asked questions
GDPR	General Data Protection Regulation
GESIS	GESIS – Leibniz Institute for the Social Sciences
GREG	Generalized regression estimators
HDD	Hard Disk Drive
IALS	International Adult Literacy Survey
iCMS	International Case Management System
ICT	Information and Communications Technology
IDB	International Database Analyzer
IEA	International Association for the Evaluation of Educational Achievement
IRT	Item Response Theory
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
ISIC	International Standard Industrial Classification of All Economic Activities
MOS	Measure of size
nCMS	National Case Management System
NR	Nonresponse
NRB	Nonresponse bias
NRBA	Nonresponse bias analyses
OECD	Organization for Economic Co-operation and Development
PDS	PIAAC Delivery System
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment
PSU	Primary sampling unit
PUF	Public Use File
ROA	Research Centre for Education and the Labour Market

RoB	Range of bias
RR	Response rate
SRS	Simple random sampling
SUF	Scientific Use File